Design and Evaluation of Context Specific User Interface for CFIT Risk Mitigation in Mountainous Operations

by

Aayush Bhattarai

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 8, 2026

Keywords: CFIT, cognitive workload, HRV,
NASA-TLX, eye-tracking, user interface

Approved by

Nicoletta Fala, Chair, Assistant Professor of Aerospace Engineering
Cheryl Seals, Professor of Computer Science and Software Engineering
Missie Smith, Assistant Professor of Industrial and Systems Engineering
Joaquim Neto Dias, Lecturer of Aerospace Engineering

Abstract

Pilots operate in information dense environments that require continuous instrument scanning, communication, navigation, weather monitoring, traffic awareness, and external visual assessment. These demands become more challenging during mountainous operations, where terrain proximity, rapidly changing hazards, or competing sources of information increase workload and reduce decision time. Since human working memory capacity cannot increase to accommodate high workload flight conditions, flight deck interfaces should reduce unnecessary information processing demand and support timely situation assessment to improve pilot performance.

This dissertation develops and evaluates a context-specific interface framework that can support different aviation operations by separating stable display architecture from operation-specific information content. The framework first establishes a fixed-zone interface architecture from human factors principles, aviation display guidance, military display standards, and visual search and decluttering literature. I used mountainous operations to validate the generic interface, as operations at high altitude and around mountainous terrain often result in Controlled Flight Into Terrain (CFIT) accidents. I used complex network analysis and information from the NTSB accident database to identify dominant information categories associated with mountainous operations/CFIT accidents. Decision making, altitude control, aircraft control, and environmental monitoring were central factors in the accident network. These factors guided the population of the generic interface placeholder with CFIT-relevant information, which included terrain relative cues, aircraft performance information, route

context, and alerts. The resulting tailored prototype was implemented in a flight simulation environment (experimental group) and evaluated against a baseline cockpit condition (control group) in two human subjects experiments.

The evaluation used subjective, physiological, operational, and visual attention measures including NASA-TLX, SART, heart rate variability, pupil dilation, flight path deviation, dwell time, and fixation measures. The tailored interface group reported improvements in perceived performance and situation understanding, along with reduced frustration. HRV results showed lower task lnRMSSD in the tailored interface group, suggesting task related engagement. Flight path deviation suggested tighter path control in the tailored interface group, although the flight path differences did not reach statistical significance. Visual attention showed that both groups allocated more attention to attitude information, while the tailored interface groups distributed less attention to some conventional instrument tape and also attended the interface.

Overall, this dissertation contributes to a human factors-based, accident-informed method for augmenting modular flight deck interfaces. This framework can adapt to different operational contexts by having a fixed zoned layout while using context specific risk analysis to determine which information should receive display priority. The CFIT case study demonstrates one application of the framework and provides initial evidence for its usefulness in supporting pilot situation assessment during high-risk operations.

Artificial Intelligence (AI) Disclosure Statement

In the preparation of this dissertation, the following Artificial Intelligence (AI) tools were used: ChatGPT, Claude, Grammarly. These tools were used primarily to troubleshoot and improve python scripts used in the analysis and in the writing process as editorial tools helping to reformulate certain sentences to improve clarity and structure. The author acknowledges full responsibility for the intellectual content of this work and has ensured that all AI-assisted sections have been reviewed and revised for accuracy and appropriate academic style. The AI-generated content was reviewed and validated for relevance, appropriateness, and accuracy before incorporation into the final document to maintain scholarly integrity of this research.

Digital Accessibility Disclosure Statement

In the preparation of this dissertation, the following digital accessibility tools were used to ensure this document complies with federal requirements: Adobe Acrobat Pro's Accessibility Checker and Microsoft Word's built in Accessibility Checker for verification. The author acknowledges full responsibility for the intellectual content of this work and has made a good faith effort to comply with digital accessibility requirements in publishing, wherein the nature of the content does not significantly change in order to do so. Furthermore, all content has been reviewed and revised to meet these requirements prior to final publication.

Acknowledgements

I dedicate this dissertation to my wife, Ashmita, my parents; and my sister and brother-in-law, whose unwavering support sustained me through every high and low.

This work would not have been possible without the guidance of my supervisor, Dr. Nicoletta Fala, and the thoughtful feedback of my committee members. I am especially grateful to ex-students cum friends Srijan, Ganesh and Sandip for always making time to offer timely and encouraging advice. I deeply cherish the moments I shared with my labmates, Luis, Niklas and Mijan, whose critical feedback and support helped shape this dissertation. I also thank Josh, Sania, and Hudson for their careful review of my work.

Finally, I dedicate this dissertation to everyone who supported, encouraged, and believed in me throughout this journey including those I may have unintentionally missed by name.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AFM Aircraft Flight Manual

AHMIs Adaptive Human Machine Interfaces

API Application Program Interface

ATC Air Traffic Controllers

CFIT Controlled Flight into Terrain

CREAM Cognitive Reliability and Error Analysis

ECG Electrocardiogram

EGPWS Enhanced Ground Proximity Warning System

FAA Federal Aviation Administration

FMEA Failure Mode and Effect Analysis

FMS Flight Management Systems

GPS Global Positioning Systems

HFDS Human Factors Design Standards

HMI Human Machine Interface

HRV Heart Rate Variability

IAF Initial Approach Fix

IATA International Air Transport Association

ICAO International Civil Aviation organization

IFR Instrument Flight Rules

IMC Instrument Meteorological Conditions

IQR Inter-Quartile Range

MAD Median Absolute Deviation

MDA Minimum Decision Altitude

NASA TLX NASA Task Load Index

PEAM Pilot Error and Analysis Method

PPNM Pixels-per-nautical-mile

QAR Quick Access Recorder

RNAV Area Navigation

ROC Rate of Climb

SA Situational Awareness

SART Situational Awareness Rating Technique

SCF Systems Component Failure

SDK Software Developer Kit

SHERPA Systmeatic Human Error Reduction and Prediction Approach

SOP Standard Operating Procedure

SOPs Standard Operating Procedures

STAMP Systems theoretic accident modelling and process

UDP User Datagram Protocol

VFR Visual Flight Rule

VR Virtual Reality

XPC XplaneConnect

1 Introduction

In aviation, pilot performance is fundamentally constrained by the limits of human cognition, specifically the capacity to manage and prioritize diverse information from instruments and environments (Colvin et al., 2009; Niermann, 2015). Despite advances in avionics, information-processing errors still account for approximately 78% of aviation incidents (NTSB, n.d.). Complex multitasking environments, such as flight operations in mountainous terrain, in which dynamic weather, reduced visibility, and terrain-induced visual illusions create high workloads and error-prone, reactive decision-making, further intensify the challenges of information processing. This strain often leads to a flawed mental model, defined here as an inaccurate representation of the aircraft state, the surrounding terrain, and the projected flight path (Endsley, 1995; Norman, 2014). As a result, the pilot overlooks or misinterprets critical data, leading to an incident or, in the absence of a sufficient recovery margin, an accident.

To address these issues, this research proposes a generic interface framework for complex multitasking environments, where the overall layout of the interface remains the same across different types of flight operations. This framework defines fixed regions on the display for critical, primary, and secondary information based on established human factors principles and Federal Aviation Administration (FAA) guidelines. What changes with the mission is not the layout itself, but the information placed within the regions. For example, during mountainous operations, task-critical information, such as terrain awareness and high-density altitude, is placed in the most prominent areas of the display. In contrast, less critical information is moved to secondary or user-configurable regions. In this work, I used mountainous flight operations as a case study to demonstrate and validate the approach.

Within my framework, I identify critical nodes to populate fixed high-salience zones and use subjective (NASA TLX) and objective metrics (eye-tracking, HRV) to estimate workload reductions in flight simulations and to evaluate the interface. The framework's modular architecture enables scalability to other domains, such as emergency response and autonomous systems. This work introduces a generalizable paradigm for cognitive overload. It does so by first using a computational method (complex network theory) to determine which factors are most critical in each context, and then strategically designing the interface to make that information highly noticeable (salient) to the pilot.

The primary contribution of this research is a generalizable design framework for proactive, context-aware cockpit interfaces. This framework is instantiated as a fixed-layout interface prototype solely for empirical validation, using Controlled Flight into Terrain (CFIT) in mountainous terrain as a representative high-risk case study.

1.1 Background

Pilots operate under a loop of perception-memory-decision-actions, a cycle that depicts all cockpit operations. Information processing refers to how people perceive, process, store, and retrieve information (Wickens & Carswell, 2021). Figure 1 illustrates the fundamental information processing cycle that underlies human perception, memory, decision-making, and action.

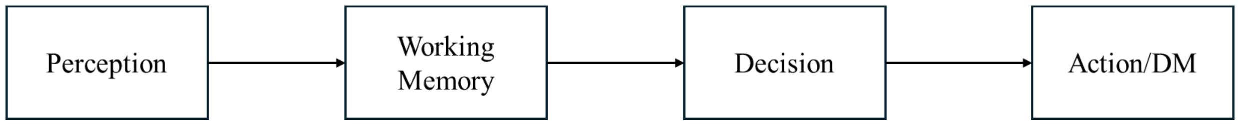

Figure 1 Fundamental model of the human information processing cycle (adapted from Wickens & Carswell, 2021)

The information processing cycle is driven by the pilots' continuous interaction with the environment. The loop begins with perception, in which sensory receptors collect raw stimuli, and attention selectively filters inputs for further processing. The chosen data then enters working memory, a limited-capacity store that holds five to seven "chunks" of information at once, where the brain actively interprets this input by recognizing patterns and relating it to existing knowledge to assign meaning. Decision-making evaluates the interpreted inputs against the pilot's immediate objective *(e.g., avoiding the cloud or maintaining a certain speed)* and existing knowledge from training and experience. Finally, the action stage is the execution of the chosen physical or mental response, which can then generate new sensory input, restarting the continuous cycle (Wickens & Carswell, 2021).

In a demanding environment such as the flight deck, hundreds of these micro-cycles unfold every minute as pilots scan instruments, interpret readings, and act, all while external conditions evolve. Each shift of attention incurs a time cost (Sewell et al., 2016), and the overall rate and volume of incoming data directly drive mental workload (P. Wang et al., 2018; Xiao et al., 2015). When a pilot overloads this cycle, attentional tunneling and working memory overload can occur. Attentional tunneling causes a pilot to fixate on one task while ignoring other vital cues, whereas working memory overload occurs when incoming information exceeds the brain's immediate processing capacity.

The overload conditions do not arise from a single source. Vidal et al. (2015) classify the factors that affect human information-processing efficiency into two broad categories: intrinsic and extrinsic factors. Intrinsic factors originate within the operator and include prior information, individual strategies, attention allocation, and error monitoring ability. For example, a pilot may expect to see the runway after passing the final approach fix, continue searching visually despite deteriorating visibility, fixate on airspeed while missing vertical path deviation, or fail to detect that the aircraft has descended below the required altitude. In contrast, extrinsic factors arise from task and environmental demands, including high stimulus volumes, decision complexity, motor-program demands, and the absence of clear predictive cues (Vidal et al., 2015). For example, mountainous flight may require the pilot to process terrain, weather, altitude, climb performance, airspeed, and navigation information simultaneously while responding to Air Traffic Control (ATC) instructions and maintaining the aircraft in turbulence. When these intrinsic and extrinsic factors disrupt information processing, they can contribute to cognitive failures. These failures play a major role in aviation accidents, particularly in high-workload environments where pilots must integrate multiple information sources under time pressure (De Almeida et al., 2026; Taheri Gorji et al., 2023). These include psychological stress (Crump et al., 2021), cognitive errors (Souvestre et al., 2008), inadequate training, procedural ambiguities (Shahriari & Aydin, 2018), environmental factors (Gumus & Saylam, 2023), design-induced issues (Wei et al., 2013), latent conditions (Erjavac et al., 2018), and team dynamics (M. Müller, 2004). The consequences range from runway excursions to catastrophic CFIT in mountainous regions. For complex multi-tasking flights, two drivers are especially dominant: the inherent complexity of the operational environment (Dill & Young, 2015; Wong et al., 2006; Young et al., 2013) and limitations in flight deck interfaces (Curtis et al., 2010; Doyon-Poulin et al., 2014; Janetzko & Kacem, 2024; Rudi et al., 2020; Senol, 2020).

A complex multitasking environment requires operators to manage multiple interdependent tasks that demand significant cognitive resources (Chérif et al., 2018; Devlin et al., 2020). Such environments expose operators to dynamic workload, increased cognitive load, and sustained task switching over extended periods (Devlin et al., 2020; Himi et al., 2019; Logie et al., 2010a). For pilots, this poses a major information-processing challenge because they must manage concurrent tasks, some of which involve continuous monitoring activities with no clear endpoints. These monitoring tasks are more cognitively demanding than discrete, goal-complete tasks because they require constant vigilance and cannot be entirely dismissed from working memory (Logie et al., 2010b). Aviation is a representative example of a complex multitasking environment. In a single flight segment, a pilot monitors flight instruments, communicates with ATC, manages automation and flight controls, and continuously updates their mental model of the aircraft state (Devlin et al., 2020; Gaetan et al., 2015; Stasch & Mack, 2024; U.S. Army Aviation Development Directorate et al., 2018). Empirical studies show that simultaneous task execution often overwhelms pilots, leading to task saturation, in which task demands exceed available cognitive resources (Hofer et al., 2024; Zheng et al., 2015). These multitasking demands are more pronounced in mountainous terrain operations than in routine, non-terrain-constrained flight operations (Gevrek et al., 2026; Rodriguez-Paras et al., 2021).

Flight in high-mountain regions introduces unique hazards, including rapidly changing weather, reduced visibility, terrain-induced visual illusions, severe localized gusts, and mountain waves that defy synoptic forecasts (ICAO, 1991). Winds that flow perpendicular to mountain ridges can create rotor patterns on the lee side of the terrain. These rotors produce intense horizontal airflow and can also produce sudden downdrafts greater than 1500 ft/min, which can exceed the climb capability of many piston-engine aircraft (Dobson & Campbell, 2014). Historically, aviation has relied on operational safeguards such as a checklist (Anton, 2022; S. Müller & Patel, 2012) and Standard Operating Procedures (SOPs) (H. Liu et al., 2018; Sumwalt III, 2004) to reduce routine human information processing errors. These safeguards improve procedural consistency, but they remain reactive because pilots must recall, consult, or execute them after a task demand or hazard has emerged. Human error analysis methods, such as the Systematic Human Error Reduction and Prediction Approach (SHERPA) (Stanton et al., 2006), Cognitive Reliability and Error Analysis Method (CREAM) (Felice et al., 2013), and Pilot Error and Analysis Method (PEAM) (Wei et al., 2013) help identify where and why errors might occur. However, these methods require significant expert involvement to decompose tasks and validate findings. More importantly, these reactive strategies and analytical tools, while valuable, do not provide pilots with proactive in-cockpit solutions for managing cognitive load as the situation unfolds.

In addition to the reactive nature of many safety interventions, current cockpit interfaces present another limitation, i.e., they are typically standardized and fixed. Regulatory guidelines, such as those issued by the FAA, help ensure consistency in element placement and labeling. Certification and standardization practices often result in fixed display layouts that cannot prioritize the most critical information as operational context changes.

To overcome this static limitation, this research develops a fixed-layout, context-sensitive interface framework that tailors information to the operational context. This framework follows a three-step process. First, it identifies mission-critical information using a data-driven analysis of historical accident data. In this study, I use complex network theory to analyze National Transportation Safety Board (NTSB) CFIT accident findings, where causal and contributing factors are modeled as nodes and their co-occurrences as edges. Centrality metrics, including degree, weighted degree, and betweenness centrality, identify factors that occupy structurally important positions in the accident network, such as pilot decision-making, altitude management, aircraft control, and environmental monitoring. Second, the framework maps these high-priority factors into predefined display regions based on their operational function and required salience. For example, urgent performance margin warnings appear in the header region, terrain and flight path information appear in the main-content regions, upcoming waypoint and performance details appear in the supplementary panel, and configuration controls remain in the navigation sidebar. Third, the resulting interface is evaluated through human-in-the-loop flight simulation using subjective, physiological, visual-attention, and flight-performance measures. This approach ensures that the interface preserves spatial stability while reducing the pilots' need to search across multiple displays and manually integrate dispersed information during high-workload conditions. Rather than relying on visual redesign alone, the framework uses an accident-derived network centrality to determine which information should receive higher salience in a given operational context.

This analysis identifies the critical parameters such as terrain mapping, pressure and density altitudes, and advisory information, and places them in high-salience display zones. This data-driven step is necessary because Adaptive Human-Machine Interfaces (AHMIs) still need a defensible rule for deciding which information should receive priority. Neuroadaptive systems can estimate cognitive state in real time, but workload sensing alone does not determine which mission-specific cues should appear in the most salient regions (Wen et al., 2025; Xenos et al., 2024). The study addresses this limitation by using accident-derived network centrality to define information priority before the operational risk escalates. The interface then applies a fixed-layout adaptation strategy; the spatial zones remain constant, while the displayed content and visual emphasis change according to the mission context. This strategy reduces the risk of mode confusion, in which pilots lose track of the system's current display state, and change blindness, in which pilots fail to notice important display changes during high-workload operations (Lutnyk et al., 2025; Wickens & Carswell, 2021).

1.2 Problem Statement

Information processing errors contribute significantly to aviation incidents. This work focuses on developing a generalizable design framework rather than a single interface instance, and it is validated through a prototype implemented in a representative high-risk operational context. Complex multitasking environments intensify information processing demands, and mountainous flight operations represent one such environment. These regions have their own mesoscale weather systems.

Dynamic weather, reduced visibility, terrain-induced visual illusions, and potential physiological effects such as hypoxia create high-workload, error-prone, reactive decision-making scenarios (Arthur Iii et al., 2006; Madhavan & Lacson, 2006; Oniscenko et al., 2024). This strain can lead to flawed mental representations of situational reality, in which the pilot overlooks or misinterprets critical data, potentially resulting in an accident.

Adaptive and proactive cockpit interfaces may help reduce aviation accidents associated with information-processing errors, particularly CFIT accidents. This research develops and empirically validates, through simulation-based human-performance evaluation, a generic interface framework. This framework is a structured design methodology for organizing, prioritizing, and presenting information on the flight deck, particularly in complex multitasking environments. This framework aims to mitigate information-processing errors by integrating static placeholders grounded in established human-factors principles and FAA guidelines, information on task-critical nodes, and user-configurable regions for secondary data streams, thereby enabling tailored information presentations.

I identified task-critical nodes by analyzing historical accidents, applying complex network theory. I evaluated the effectiveness of this framework in reducing information-processing errors using human-in-the-loop simulated flight scenarios and physiological estimates of the user’s workload. The framework's modular architecture enables scalability to other safety-critical domains, such as oceanic flight, by redefining network nodes to suit the specific operational context.

1.3 Research Gap

Existing aviation safety measures are often reactive interventions that address hazards after they begin to emerge, while many flight deck interfaces rely on standardized and fixed information layouts (Holley & Miller, 2022; Lee, 2012; Reason, 2000; Vanderhaegen, 2015; H. Wang et al., 2020). These limitations indicate a critical research gap: aviation lacks a proactive, data-driven interface framework that can prioritize mission-critical information and mitigate pilot information-processing errors in complex multitasking environments, such as mountainous operations.

This gap persists because current approaches have shortcomings. Reactive tools such as checklists and training do not modify the information environment to reduce cognitive load in real time. In contrast, even advanced adaptive interfaces have not been sufficiently validated to use historical accident data to proactively prioritize information. Therefore, there is a clear need for a system that bridges the gap between human factors theory and a practical, data-driven, safety-focused application. This research aims to address this specific gap by proposing and validating a design framework implemented as an interface-based intervention.

1.4 Research Objectives

This research aims to develop and evaluate an interface framework to proactively mitigate pilot cognitive overload and enhance information processing, particularly in complex multitasking environments such as mountainous terrain. I will achieve this research aim through the following objectives:

Design, develop, and evaluate a decision-support interface framework that, informed by historical aviation accident data, strategically presents mission-critical flight parameters within predefined, high-salience display zones.

Investigate cognitive stressors and error patterns affecting pilot decision-making during mountainous approach flights using virtual reality (VR) flight simulations.

Evaluate the impact of the developed decision-support interface framework on pilot performance by assessing reductions in subjective cognitive workload (e.g., using NASA-TLX) and objective physiological stress indicators (e.g., eye-tracking metrics, Heart Rate Variability), and measuring improvement in information processing efficiency, situational awareness, and the quality and timeliness of decision-making under simulated high-workload conditions representative of mountainous flight.

Analyze the practical potential of the developed framework as an intervention to enhance flight safety in mountainous environments, comparing its proactive approach to existing reactive safety measures and the current interface technologies.

1.5 Scope

The proposed experiment uses a Cessna 172 model, limiting the direct generalizability of findings to other aircraft types or multi-engine operations. While numerous factors influence aircraft performance, this study's primary objective is to reduce information processing load for pilots during mountainous flights by displaying data identified as critical through our network analysis. Consequently, the framework does not explicitly address or account for other factors that might impact aircraft performance or pilot decision-making beyond the scope of information processing.

1.6 Dissertation Structure

I have organized this document as follows. Chapter 2 provides a literature review of information processing in aviation and offers a rationale for proactive measures. Chapter 3 details how I designed the generic interface. Chapter 4 provides the study's context for CFIT and explains how I populated the generic interface with information elements. Chapter 5 outlines the process by which the prototype was designed. Chapter 6 discusses the methods used and results from the subjective and objective data. Chapter 7 provides the conclusion and recommendations.

2 Literature Review

This section reviews the literature on information processing, cognitive load, and aviation safety to support the argument that a modular, data-driven interface framework is a necessary and logical next step for enhancing flight safety.

2.1 Human Information Processing

The flight deck represents a high-density information environment in which pilots must continuously perceive, interpret, and respond to dynamic stimuli under time pressure. Human performance in such environments depends not on the sophistication of aircraft systems alone, but on the limits of cognitive architecture that govern information processing (Colvin et al., 2009; Niermann, 2015). Any attempt to improve aviation safety must therefore begin with an understanding of how humans process information.

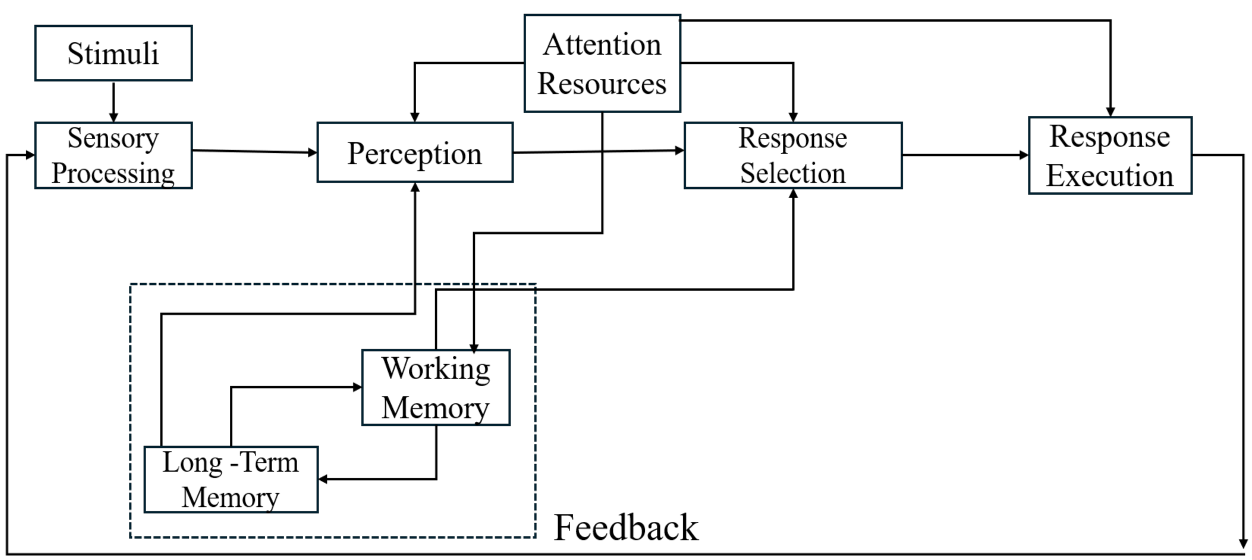

Figure 2 Human Information Processing models illustrating sensory processing, interaction between working memory and attentional resources, allocation of attentional resources, response selection, and feedback loops (adapted from Wickens et al. 2023)

Classical information processing theory conceptualizes cognition as a staged process involving sensory registration, perceptual encoding, working memory integration, response selection, and motor execution (Wickens et al., 2023). Environmental stimuli first enter sensory memory, where they persist briefly before attention filters relevant signals for further processing. Perception transforms selected stimuli into meaningful representations. Working memory then maintains and manipulates this information while retrieving relevant knowledge from long-term memory. Response selection evaluates possible actions, and motor execution implements the chosen response. Feedback from action re-enters the perceptual system, forming a closed loop (Wickens et al., 2023).

Working memory functions as the central bottleneck of the system, and the primary theoretical lens for understanding this limitation is cognitive load theory. This theory defines cognitive load as the amount of information simultaneously held and manipulated within working memory (Sweller, 1988). People can actively maintain approximately five to seven meaningful units, or “chunks” of information, at one time (Miller, 1956). Although subsequent research refined these limits, agreement remains that working memory capacity is sharply bounded (Johnson et al., 2014; Reid, 2009; Yamaguchi & Bortz, 2021).

Task complexity increases cognitive demand through several mechanisms (Zheng et al., 2015). First, complexity rises when the number of information sources increases. Multiple instruments, alerts, communication, and environmental cues compete for attention. According to the multiple resources theory, tasks that draw from overlapping perceptual and cognitive resources create interference and reduce processing efficiency (Wickens, 2008). Second, complexity increases when elements must be processed simultaneously rather than independently (Chen et al., 2023). Cognitive load theory describes this as high element interactivity, which requires integration of multiple variables within working memory (Sweller, 1988). For example, interpreting altitude alone imposes minimal demand; integrating altitude, terrain proximity, vertical speed, aircraft performance, and environmental density requires coordinated processing across several interacting variables. Third, multitasking introduces task-switching costs (S. Liu et al., 2016). Pilots must alternate between navigation, communication, monitoring, and aircraft control. Each switch requires reallocation of attentional resources and reconstruction of task context, which consumes working memory capacity and increases vulnerability to omission errors (Logie et al., 2010b). Fourth, time pressure compresses decision windows and narrows attentional focus (J. R. Kelly & Loving, 2004; Wu et al., 2022). Under high workload, attention may tunnel toward a single salient cue while suppressing competing information streams. For example, a pilot may fixate on maintaining airspeed during a high-workload approach while failing to monitor vertical path deviation, proximity to terrain, or configuration status. This phenomenon reduces situation monitoring and increases the likelihood of missing critical signals (Moehlenbrink et al., 2011).

When attentional demand exceeds available cognitive resources, a breakdown occurs within the information-processing loop. Working memory overload can delay response selection, distort interpretation of instrument data, or disrupt procedural recall (Hadar et al., 2016; Xin & Li, 2020). Attentional tunneling can cause fixation on one instrument or task while neglecting others (Moehlenbrink et al., 2011). These failures do not result from insufficient cockpit information; rather, they stem from a structural mismatch between cognitive capacity and environmental demand.

In safety-critical domains such as aviation, even minor inefficiencies in information structuring can push cognitive systems toward saturation. Therefore, interface design must align with the constraints of working memory, attentional allocation, and response selection processes. Any framework that aims to enhance aviation safety must first acknowledge these inherent cognitive limits and structure the information environment accordingly.

2.2 Situation Awareness and Mental Model

Situational Awareness (SA) describes the operator's ability to perceive environmental elements, comprehend their meaning, and project their future status within a given time horizon (Endsley, 1995). In aviation, SA determines whether a pilot correctly understands aircraft state, environmental conditions, and system constraints in time to make appropriate decisions.

Endsley’s three-level model of SA provides a structured framework. Level 1 (Perception) involves detecting relevant cues, such as altitude, airspeed, terrain alerts, and weather indications. Level 2 (Comprehension) requires integrating these cues into a coherent understanding of the system state. Level 3 (Projection) involves anticipating future system behavior, such as predicting descent profile relative to terrain or fuel state relative to route distance. Each level builds on the previous one and depends on the integrity of the underlying information-processing mechanism.

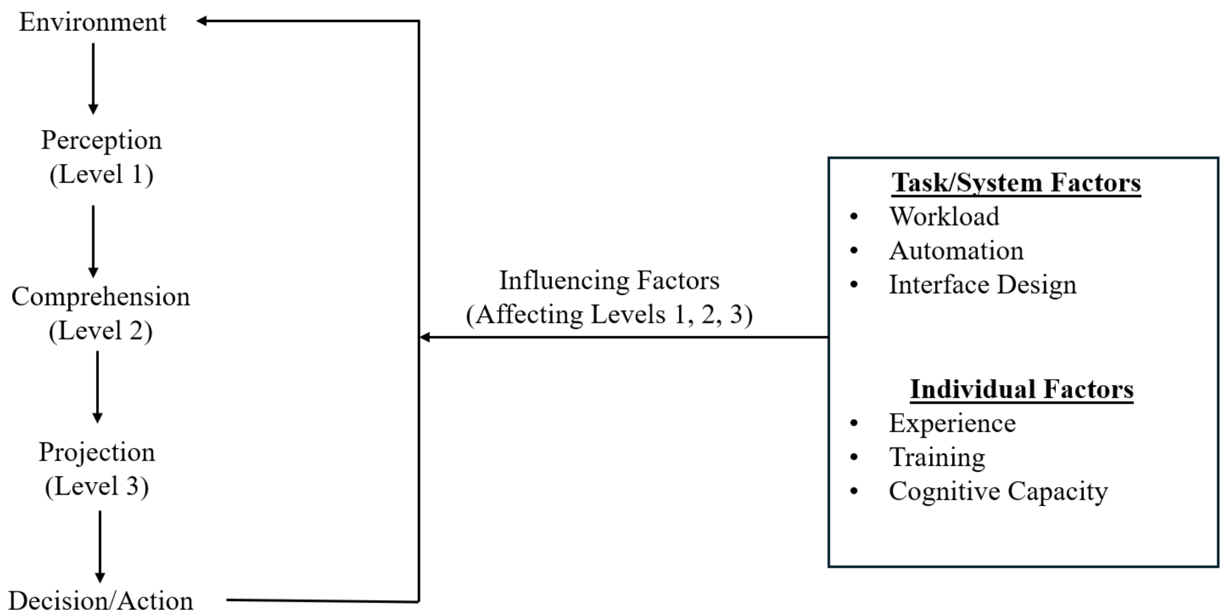

Figure 3 Hierarchical model of individual and task/system factors influencing all levels of SA (inspired by Endsley, 1995; Wickens & Carswell, 2021)

Simply presenting more data to pilots does not guarantee situational awareness. Pilots must identify the relevant cues, interpret their operational meaning, and use them to anticipate future aircraft, terrain, and environmental states. It depends on the construction and continuous updating of mental models (Zaw & Suyatinov, 2022). A mental model represents an internal representation of how a system works or behaves (Johnson-Laird, 2012). In aviation, pilots rely on mental models to interpret instrument readings, anticipate aircraft behavior, and predict environmental change (Wickens et al., 2023). Long-term memory stores procedural knowledge and system rules, while working memory integrates incoming information with these stored representations to maintain a dynamic understanding of flight state (Schlimm, 2015; Sohn & Doane, 2004).

Breakdown in SA often precedes operational error (Kalagher et al., 2021). CFIT refers to an event in which airworthy aircraft, under pilot control, unintentionally impacts terrain, water, or an obstacle because the crew does not recognize the hazard in time to avoid it. CFIT events, unstable approaches, and altitude deviation are often associated with degraded perception, incomplete comprehension, or failed projection rather than a lack of available data (Kalagher et al., 2021). The cockpit contains all the necessary information, yet failure to structure and integrate it leads to inaccurate mental models and flawed decisions (Lazaro, Kang, et al., 2021).

Decision quality depends directly on the integrity of situational awareness (Botega et al., 2019). Accurate perception enables correct problem identification, effective comprehension supports prioritization among competing demands, and reliable projection allows proactive risk management rather than last-minute correction (Endsley, 1995). Therefore, interface design must support not only information visibility but also the structured development of perception, comprehension, and projection. Effective design must reduce unnecessary cognitive burden and promote the construction of coherent mental models.

2.3 Multitasking, Workload, and Attention Allocation

Modern flight operations require pilots to manage multiple concurrent tasks, including aircraft control, navigation, communication, automation monitoring, terrain awareness, weather monitoring, and external visual scanning (Devlin et al., 2020; Gaetan et al., 2015; Stasch & Mack, 2024). These tasks never occur sequentially; they compete for shared cognitive resources. Multitasking in aviation, therefore, depends on how attentional resources are allocated across parallel demands (Xie et al., 2025). To understand what drives these allocation demands, it is useful to examine the nature of the tasks themselves through the lens of Cognitive Load Theory, which distinguishes among intrinsic, extraneous, and germane load (Sweller, 1988).

Intrinsic load arises from the task's inherent complexity. For example, executing a non-precision approach in mountainous terrain involves integrating terrain awareness, aircraft performance, and weather conditions. Extraneous load arises from the way information is presented. Poorly organized displays, excessive visual clutter, or fragmented data sources increase extraneous demand without improving task performance. Germane load refers to the mental effort a person uses to build useful knowledge patterns, or schemas, that help them to understand and perform a task more effectively (Jordan et al., 2020). For example, a pilot develops a schema when they learn to connect altitude, terrain clearance, vertical speed, and aircraft performance into a single understanding of whether the aircraft can safely clear the arising terrain. In aviation, extraneous load is particularly critical because it consumes limited working memory resources that could otherwise support comprehension and projection (Loft et al., 2023). One significant contributor to this extraneous load is the cognitive cost of shifting attention between concurrent tasks. When pilots switch between tasks, the resulting interference is particularly pronounced when those tasks compete for the same perceptual and cognitive resources, a central concern in understanding the demands of multitasking.

Multitasking introduces an additional cognitive cost through task switching (Bandodkar & Singh, 2014). Each shift of attention requires disengagement from one task and re-engagement with another. This process incurs switching costs, including temporary performance degradation and an increased likelihood of errors. When multiple tasks require overlapping perceptual modalities or cognitive processes, interference increases according to the Multiple Resource Theory (Wickens, 2008). For example, visual scanning of flight instruments while monitoring terrain displays and reading textual alerts draws heavily from shared visual resources, increasing competition within the perceptual channel. This intersection between perceptual channels, cognitive processing stages, and response types is formally described by Multiple Resource Theory.

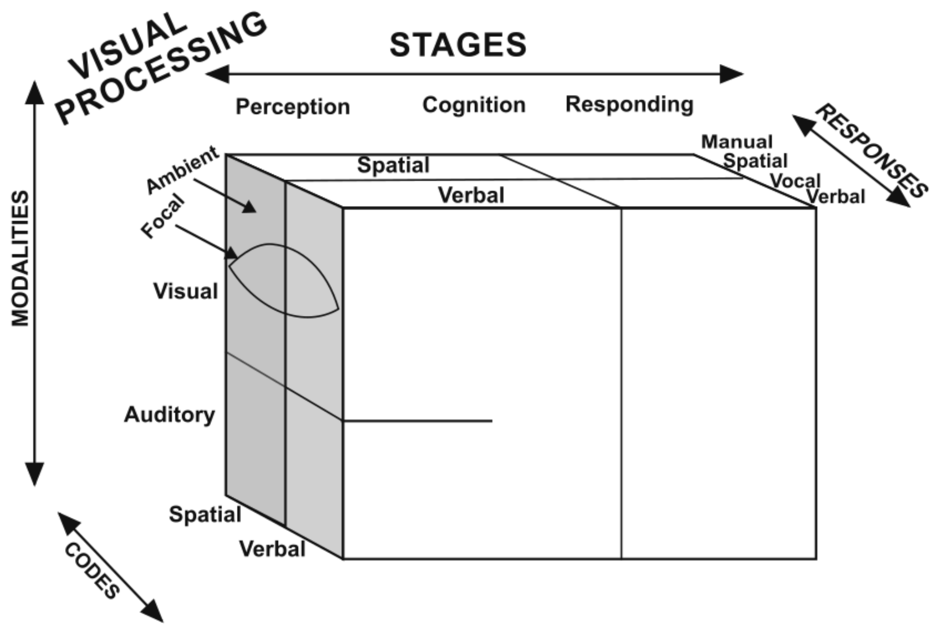

Figure 4 The Multiple Resource Theory model depicting attentional resources across processing stages, perceptual modalities, processing codes, and response types (Wickens, 2008).

Attention allocation determines which information enters working memory and which information remains unattended (Sherman & Turk-Browne, 2024). In dense cockpit environments, attentional resources must be distributed across primary flight instruments, navigation displays, automation panels, and external visual cues. As display density increases, visual search time increases, and the probability of missing critical information arises (Moacdieh & Sarter, 2017). When visual prominence does not match operational importance, cognitive cost increases. For example, consider a display in which weather radar imagery flashes prominently during final approach, a phase when vertical guidance is critical for a safe landing. Meanwhile, a deviation from the glideslope (the optimal descent path) appears only as a subtle needle movement. The pilot's attention is captured by the flashing weather data (high salience, low immediate relevance) while actively searching for the glideslope deviation (low salience, high immediate relevance). This search imposes cognitive cost: attention must shift from the salient distraction, visual scanning time increases, and working memory is burdened with locating and interpreting the vertical path information that should have been immediately perceptible. The pilot spends mental effort locating the glideslope data rather than using it to correct the approach.

Under elevated workload conditions, attention narrows and becomes selective. Operators may prioritize the most salient or immediately demanding task while deprioritizing monitoring functions. This narrowing reduces the breadth of situation awareness and weakens the Level 3 projection (the ability to anticipate future states) (Endsley, 1995). In time-critical environments, this shift from distributed monitoring to reactive control can occur rapidly.

Importantly, workload does not arise solely from task quantity; it emerges from the interaction between task demands and system design. A system that requires manual integration of various dispersed information elements increases the extraneous load (Pink & Newton, 2020), while a system that aligns the same information with operational relevance reduces cognitive burden (Michailovs et al., 2023). Therefore, workload management is not exclusively an operator responsibility but also a design responsibility.

Therefore, effective interface design must account for attentional limits, minimize unnecessary task switching, and structure information to reduce visual search cost. When the system fails to support efficient attention allocation, cognitive overload becomes more likely, increasing the probability of perceptual omission, incomplete comprehension, and degraded projection.

2.4 Limitations of Reactive Safety Paradigms in Aviation

Aviation safety has historically improved through the introduction of procedural safeguards, training protocols, and warning systems designed to mitigate operational risk. While these mechanisms have significantly enhanced flight safety, many operate primarily as reactive interventions, responding only after a hazardous state has begun to develop (Provan et al., 2020; Woods, 2015). These systems help pilots recover after a hazardous condition has already developed, but they do not reduce the cognitive workload that may have contributed to the hazard in the first place.

Procedural interventions, such as checklists and SOPs, are reactive strategies. These tools structure pilot actions during critical phases of flight and reduce reliance on memory during complex operations. However, their effectiveness still depends on the pilot's ability to recall, interpret, and execute procedural steps under time pressure. During high-workload conditions, attentional resources are limited, increasing the likelihood that checklist items are delayed, skipped, or incorrectly completed (Chou et al., 1996; Degani & Wiener, 1993; Parasuraman et al., 2008). As a result, procedural safeguards often address the symptoms of workload rather than its underlying causes.

Technological safety systems also largely follow a reactive paradigm. A well-known example is the Terrain Awareness Warning System (TAWS), which alerts the pilot when the aircraft trajectory poses an imminent risk of terrain collision. While TAWS has significantly reduced CFIT accidents, the system typically activates only when the aircraft has already entered a hazardous flight state. At this point, the pilot must respond rapidly to warnings while simultaneously managing aircraft control, navigation, and communication demands. Consequently, the system functions as a last-line defensive barrier rather than a proactive cognitive support mechanism (D. Kelly & Efthymiou, 2019; Singer & Dekker, 2000).

The distinction highlights the broader difference between reactive and proactive safety interventions. Reactive interventions act after hazardous conditions have begun to develop (Provan et al., 2020; Woods, 2015). For example, TAWS warns the pilot when the aircraft's trajectory already poses a risk of colliding with terrain (D. Kelly & Efthymiou, 2019; Singer & Dekker, 2000). Proactive interventions act earlier by organizing and prioritizing information before the pilot reaches a high-workload or time-critical state (Andre et al., 1991; Parasuraman et al., 2008; Wickens & Carswell, 1995). This research focuses on a proactive approach because the proposed interface aims to reduce visual search, support earlier risk interpretation, and help pilots make decisions before situations become critical.

Another limitation of many modern flight deck systems is the static nature of cockpit interfaces. Human factors research shows that display formatting, information proximity, and adaptive automation can significantly affect situation awareness and workload (Andre et al., 1991; Kaber & Endsley, 2004; Wickens & Carswell, 1995). As a result, the relative importance of information is not adjusted to match mission-specific demands. For example, parameters such as density altitude, terrain clearance, or aircraft climb performance may become critically important during high-altitude mountainous operations yet remain visually secondary within the display hierarchy. Pilots must therefore manually search for, retrieve, and integrate this information, increasing cognitive workload during already demanding flight phases.

Static alerting systems further contribute to this challenge. Alerts typically activate only when predefined thresholds are exceeded, providing warnings that may arrive too late to support anticipatory decision-making. Additionally, alerts are often presented independently of broader contextual information, forcing pilots to interpret warning signals while simultaneously reconstructing the surrounding operational situation.

Taken together, these limitations place a significant cognitive burden on pilots, particularly during high workload phases of flight. Pilots must continuously integrate dispersed information, detect emerging hazards, and anticipate future system states while managing multiple concurrent tasks. When the system does not actively support this information-integration process, cognitive overload is more likely, increasing the risk of perceptual omission, incomplete comprehension, and degraded situational awareness.

2.5 Static Cockpit Interfaces and Information Hierarchy Constraints

Modern flight decks present large volumes of information through integrated avionics displays, such as electronic flight instrument systems and glass cockpit architectures. While these systems provide pilots with unprecedented access to flight, navigation, and environmental data, the underlying display structures are often based on fixed interface architectures in which the spatial arrangement and visual hierarchy of information remain largely unchanged across different operational contexts (Holley & Miller, 2022; H. Wang et al., 2020). These static layouts prioritize consistency and standardization, but they do not dynamically adjust information salience to meet the operational demands of a specific mission.

In static cockpit interfaces, the relative prominence of displayed information is typically determined during the design phase and remains constant regardless of changing environmental or operational conditions. As a result, information that may be operationally critical during a particular phase of flight may not receive sufficient visual priority within the display hierarchy. Pilots must therefore actively search for relevant parameters and mentally integrate them with other available cues, thereby increasing cognitive workload during already demanding tasks.

One consequence of static display architectures is the lack of contextual prioritization. Different flight scenarios impose different information requirements. For example, terrain clearance, density altitude, and aircraft climb performance may become particularly important during high-altitude departures or approaches in mountainous regions. However, conventional cockpit displays often retain the same information hierarchy used during routine en route operations. When operationally relevant parameters remain visually secondary, pilots must rely on manual scanning and memory-based integration to extract the information required for safe decision-making.

Static interfaces can also contribute to information clutter and increased visual search burden. As avionics systems incorporate additional capabilities, displays frequently accumulate multiple layers of information, including navigation overlays, weather radar imagery, terrain mapping, and traffic advisories. While each element may provide useful situational information, the simultaneous presence of many display elements can increase visual complexity. Increased display density has been shown to lengthen visual search time and increase the likelihood of missing critical cues during time-sensitive phases of flight (Moacdieh & Sarter, 2017).

Another important limitation of static display systems is the misalignment between mission risk and display salience. Information that carries high operational risk may not always appear with corresponding visual emphasis. For instance, parameters related to terrain proximity or vertical flight-path deviation may appear as small indicators within the primary flight display, while other, less time-critical information may occupy larger or more visually prominent regions of the interface. When display salience does not align with operational risk, pilots must allocate additional attentional resources to identify and interpret critical cues.

These limitations illustrate how static cockpit interfaces can increase cognitive workload by requiring pilots to continuously filter and prioritize information. Rather than supporting the pilot's cognitive processes, the interface can unintentionally increase attentional demand during high-workload phases of flight. As a result, effective cockpit design must consider not only the availability of information but also how that information is organized, prioritized, and presented in relation to the operational context.

Understanding these interface constraints provides a foundation for examining how human factors design principles attempt to manage information salience and support pilot decision-making. The next section reviews human-factors standards and interface design principles that guide the organization of cockpit information and the prioritization of alerts.

2.6 Interface Design and Decision Support in Aviation

Human Machine Interface (HMI) design in aviation is guided by established human factors principles to ensure cockpit information supports safe and efficient decision-making. Regulatory and military standards, such as the Federal Aviation Administration Human Factors Design Standards (HFDS), Electronic Flight Displays, Human Factors Design Guidelines for Multifunction Displays, and MIL-STD-1472, MIL-STD-783D, MIL-STD-411, provide guidelines for display layout, information & alert prioritization, and workload management in complex operational environments (DOD, 1984, 1991, 1996; FAA, 2014, 2016; Mejdal et al., 2001). These standards emphasize that cockpit interfaces must present information in ways that align with human perceptual and cognitive capabilities.

One important principle within these guidelines is alert prioritization (DOD, 1991). Alerting systems are typically organized into hierarchical categories that reflect the urgency of the underlying hazard. For example, many modern avionics systems classify alerts into warning, caution, and advisory levels, each associated with distinct visual and auditory characteristics. Warnings indicate conditions requiring immediate corrective action, caution signals situations requiring timely pilot attention, and advisories provide informational messages that support situational awareness. Proper prioritization ensures that the most time-critical information receives the greatest perceptual salience (meaning it stands out visually), reducing the likelihood that urgent alerts are overlooked during high-workload situations (Hsueh & Chen, 2025; Y. Li et al., 2018).

Another key design concept is the organization of information grouping and display (Mejdal et al., 2001). Human Factors standards recommend that related information be presented together in ways that support efficient perceptual processing and reduce visual search time (M. Yang et al., 2024). Grouping flight parameters, navigation data, and system status indicators into coherent display regions allows the pilot to interpret information more quickly and reduces the cognitive effort required to integrate dispersed data elements (FAA, 2016). When information is organized according to functional relationship rather than arbitrary spatial placement, a pilot can more easily form accurate mental models of the system state.

Closely related to information grouping is the concept of visual salience management (DOD, 1996; FAA, 2016). Effective interface design uses visual features such as color, contrast, size, and motion to direct the pilot's attention to the most important information (Draeger, 2009; Treisman & Gelade, 1980). By aligning visual prominence with operational relevance, interface designers can reduce unnecessary visual search and support faster recognition of critical cues. Conversely, when visual salience does not align with operational importance, pilots must rely on deliberate scanning strategies to locate essential information, thereby increasing attentional demand (Bellenkes et al., 1997).

Beyond simple information presentation, modern cockpit interfaces increasingly incorporate decision-support functions that assist pilots in interpreting complex operational data (Menig et al., 2025; Ramos et al., 2023). Decision-support systems may provide predictive information, highlight emerging hazards, or integrate multiple data sources into simplified visual representations that support faster comprehension

(Bhattacharya, 2025; Salvador et al., 2022). Examples include flight management systems (FMS) that compute optimal trajectories, terrain awareness systems that provide predictive alerts, and energy management cues, such as approach path indicators, that assist pilots in maintaining stable approach profiles.

Despite these advances, many cockpit decision-support tools remain limited in their ability to dynamically adjust information presentation in response to changing operational contexts (Mishra et al., 2017; Torres et al., 2015). Most systems rely on predefined thresholds or static display layouts rather than continuously adapting the information hierarchy to reflect mission-specific risks (Godfroy-Cooper et al., 2022; Letsu-Dake et al., 2014). As a result, pilots often remain responsible for integrating dispersed information sources and determining which parameters require the greatest attention at any given moment (Hofer et al., 2024; Meier, 2021).

These limitations suggest that further improvements in cockpit interface design may require approaches that move beyond static display rules toward more adaptive information management strategies. The following section, therefore, reviews research on adaptive and context-sensitive interface systems that attempt to adjust information presentation based on operational conditions and pilot workload.

2.7 Adaptive and Context-Sensitive Interface Approaches

Researchers have increasingly explored adaptive HMI to better align cockpit information presentation with dynamic operational demands (Moens et al., 2026; Xenos et al., 2024). Adaptive interfaces modify elements of the display environment based on changes in system state, environmental conditions, or operator workload (Rothrock et al., 2002). Rather than presenting information through a fixed hierarchy, these systems adjust visual salience, display content, or group information in response to contextual factors. In aviation environments characterized by rapidly changing operational conditions, adaptive interfaces have the potential to reduce cognitive workload and support more effective information processing (Holley & Miller, 2022; Wen et al., 2025).

One approach within this domain involves context-sensitive displays, which modify the visibility or priority of information based on the current phase of flight or mission context (Burian et al., 2023; Lauren et al., 2018; Lutnyk et al., 2025). For example, during approach and landing phases, systems may highlight vertical guidance information, terrain proximity cues, or runway alignment indicators. During en-route cruise, the interface may emphasize navigation performance, fuel management, or weather monitoring. By dynamically prioritizing information relevant to the current operational task, context-sensitive interfaces attempt to reduce unnecessary visual search and improve pilot situational awareness (X. Yuan et al., 2025).

Despite these advantages, adaptive cockpit interfaces also present several challenges. One major concern involves predictability and operator trust. If display elements change unexpectedly or the system modifies the presentation of information in ways not transparent to the pilot, operators may experience confusion or reduced trust in the interface (Etherington et al., 2020; Gajos et al., 2008; D. Liu et al., 2012). This issue, often described as an automation surprise, occurs when automated system behavior differs from the pilot’s expectations (Sarter & Woods, 1997). Human factors research has shown that maintaining consistent mental models of automation behavior is essential for safe human–automation interaction (Parasuraman et al., 2000).

Another challenge involves determining which information should be prioritized under operational conditions. Many adaptive interface studies rely on heuristic design decisions or limited task analyses to determine display changes (Bachmann et al., 2006; Botello et al., 2006; Saffell et al., 2011). However, these approaches do not always incorporate empirical accident data or systematic risk modeling when determining which information elements should receive higher visual salience (Xiao et al., 2025). As a result, adaptive display systems may adjust interface elements without fully accounting for the underlying risk structures associated with specific flight operations.

In addition, empirical validation of adaptive cockpit interfaces remains relatively limited (Arthur et al., 2018; Cahill et al., 2016; Frank & Aaseng, 2016; Lauren et al., 2018; Lim et al., 2018; D. Liu et al., 2012; López-Contreras, 2025). While simulator studies have demonstrated improvements in workload and attention allocation under controlled experimental conditions, fewer studies have evaluated adaptive systems within operationally complex environments or across diverse mission contexts (Ding et al., 2025; Henneberry et al., 2026). The absence of large-scale validation studies makes it difficult to determine how adaptive interface concepts generalize across different aircraft types, operational scenarios, and pilot populations.

These limitations suggest that further progress in adaptive cockpit interface design may require stronger integration between human factors design principles, operational risk modeling, and empirical validation frameworks. In particular, identifying which information elements should receive priority within adaptive displays may benefit from a systematic analysis of accident data and risk propagation patterns within aviation systems.

The next section explores data-driven risk modeling methods, including complex network techniques for analyzing accident causes and risk flows in aviation systems. These methods lay the groundwork for identifying key risk factors to help prioritize information in adaptive cockpit interfaces.

2.8 Data-Driven Risk Modeling and Complex Network Approaches

Traditional aviation safety analysis has historically relied on linear accident models, which represent accidents as a sequence of cause-and-effect events (Heinrich, 1941; Katsakiori et al., 2009). Early frameworks, such as Heinrich's accident triangle, and later fault tree and event-tree method, assume that accidents occur through identifiable chains of failures progressing from initial hazards to final outcomes (Ericson, 2005; Heinrich, 1941; Katsakiori et al., 2009; Ruijters & Stoelinga, 2015). These approaches have been widely used in aviation safety management because they provide structured methods for identifying failure points and implementing corrective actions (Salvendy & Karwowski, 2021).

However, many researchers have argued that complex socio-technical systems, such as aviation operations, cannot be fully explained by purely linear models (Rasmussen, 1997). Aircraft accidents rarely result from a single failure or a simple sequence of events (Yilmaz, 2025). Instead, they typically emerge from interactions among multiple factors, including environmental conditions, human decision-making, aircraft system behavior, and operational constraints (Shahriari & Aydin, 2018). System-oriented accident models, such as Reason’s Swiss Cheese Model and later system safety frameworks, recognize that accidents arise from the alignment of multiple latent and active failures distributed across the operational system (Reason, 2000).

To better represent these complex interactions, recent studies have begun applying network-based approaches to accident analysis (S. Li et al., 2025; Lin et al., 2014; Yilmaz, 2025; Yue et al., 2021). In these models, accident-contributing factors are represented as nodes in a network, and the relationships between factors are represented as links connecting them. This structure allows researchers to examine how risk propagates through interconnected systems rather than through simple linear chains. Network-based accident models can identify influential nodes, clusters of interacting risk factors, and pathways through which safety events escalate (Bhattarai & Fala, 2025; J.-F. Yang et al., 2023).

Complex network analysis has been applied across several transportation safety domains to examine accident-causation patterns and the propagation of systemic risk (K. Li & Wang, 2018; S. Li et al., 2025; Lin et al., 2014; J.-F. Yang et al., 2023). By analyzing large accident databases, researchers can construct networks that reveal which factors frequently co-occur and how interactions among operational, environmental, and human factors contribute to safety outcomes (Bhattarai & Fala, 2025). These methods allow investigators to move beyond single-cause explanations and instead analyze the structural relationship among contributing and causal factors within accident systems.

Within aviation safety research, network-based approaches have been used to study accident causation in areas such as air traffic management events, general aviation accident databases, etc (Bhattarai & Fala, 2025; Yilmaz, 2025; Yue et al., 2021;

Zhang & Mahadevan, 2021). These studies demonstrate that accident factors often form interconnected clusters in which certain nodes exert disproportionate influence on overall system risk. Identifying these high-influence nodes provides insights into where safety interventions may be most effective. Relatively few studies have applied complex network methods specifically to CFIT accidents occurring in mountainous environments. Mountain operations involve a unique combination of terrain constraints, weather variability, navigation complexity, and pilot workload (Baker & Lamb, 1989; Troller et al., 2016). These conditions create tightly coupled risk structures in which failures in perception, comprehension, or decision-making can rapidly propagate into hazardous flight states (Bhattarai et al., 2025; Khatwa & Roelen, 1999).

Understanding how these contributing factors interact is essential for identifying which information elements should receive priority within the cockpit display. If certain risk factors consistently occupy central positions within accident networks, then interface design strategies may benefit from emphasizing information associated with those factors. However, existing research on adaptive interfaces rarely integrates accident-derived risk structures when determining information prioritization strategies (Huttner et al., 2024; W.-C. Li et al., 2020; Richards & Lamb, 2016; Tufano et al., 2023).

Therefore, combining complex network analysis of accident data with cockpit interface design principles may provide a more systematic basis for determining which information elements should receive the highest salience during specific operational contexts. Such an approach allows interface design to be informed not only by human cognitive limitations but also by empirically derived risk relationships within aviation accident systems.

2.9 Measuring Information Processing and Cognitive Load

Assessing cognitive workload and information processing in complex operational environments requires a multi-dimensional approach. No single metric fully captures the underlying cognitive state of an operator. Subjective, physiological, and performance-based measures each provide partial but complementary insights into how individuals allocate attention, process information, and respond to task demands. Therefore, researchers commonly adopt a multimodal measurement framework to evaluate workload and information processing in the aviation context (J. Yuan et al., 2025).

2.9.1 Subjective Workload and Situation Awareness Assessment

Subjective workload assessment methods capture the operators' perceived cognitive demand during task execution. Among these, the NASA-TLX remains one of the most widely used tools for evaluating workload across multiple dimensions, including mental, physical, and temporal demands, effort, performance, and frustration (Cao et al., 2009; K. Virtanen et al., 2022).

In addition to workload assessment, the Situational Awareness Rating Technique (SART) measures perceived situational awareness by evaluating attentional demand, attentional supply, and understanding of the task environment (Braarud, 2021). Situational awareness refers to an operator's perception of elements in the environment, comprehension of their meaning, and projection of their future status (Endsley, 1995). SART complements workload measures by capturing how well operators believe they understand and anticipate system behavior. While NASA-TLX reflects perceived effort, SART reflects perceived awareness, allowing researchers to examine the relationship between workload and situational awareness (Braarud, 2021).

Despite their widespread use, subjective measures rely on self-reports and are typically collected after task completion. As a result, they do not capture real-time fluctuations in cognitive demand and may be influenced by memory biases, performance outcomes, or individual interpretations of rating scales (Braarud, 2021). These limitations motivate the use of additional objective measures.

2.9.2 Physiological Indicators of Cognitive Load

Physiological measures provide continuous and objective indicators of cognitive workload by capturing changes in autonomic nervous system activity and visual processing. Among these, heart rate variability (HRV) has been widely used to assess cognitive demand in an aviation setting (Soares et al., 2025; J. Yuan et al., 2025). HRV refers to the variation in time intervals between consecutive heartbeats and reflects the balance between sympathetic and parasympathetic nervous system activity (Pham et al., 2021). Under increased cognitive workload, parasympathetic activity typically decreases, leading to reduced HRV (Pereira et al., 2017). Metrics such as Low Frequency/High Frequency (LF/HF), Standard Deviation of NN Intervals (SDNN), and Root Mean Square of Successive Difference (RMSSD) serve as sensitive indicators of mental effort and stress during task execution (Pham et al., 2021).

Pupillometry offers another well-established measure of cognitive load. Task-evoked pupillary responses reliably reflect the mental effort required to perform cognitive tasks (Xenos et al., 2024). Changes in pupil diameter correlate with cognitive effort; increased workload leads to pupil dilation. In visually demanding environments, such as cockpit operations, pupil responses provide insight into moment-to-moment variations in information-processing demand (Stasch & Mack, 2024). Eye-tracking systems further enable the analysis of gaze behavior, fixation patterns, and visual attention allocation across cockpit displays (Xenos et al., 2024).

Although physiological measures provide high temporal resolution, they can be influenced by external factors such as lighting conditions, emotional state, and individual variability. Consequently, physiological data should be interpreted alongside other measures to ensure robust conclusions.

2.9.3 Performance-Based Indicators

Performance-based measures assess how cognitive workload manifests in task execution (Tamilselvan et al., 2023). In aviation contexts, deviations from desired flight parameters serve as a key indicator of information-processing limitations. When cognitive demand exceeds available attentional resources, pilots may fail to maintain precise control of aircraft states. In Instrument Flight Rules (IFR) environments, a pilot's primary task is to precisely control the aircraft to follow an assigned route and altitude, making deviations a clear sign of task saturation.

Altitude deviation and lateral path deviation serve as key indicators of performance degradation (Tamilselvan et al., 2023). Deviations from the intended flight path reflect breakdowns in perception, comprehension, or decision-making processes. For example, failure to maintain a stable descent profile may indicate reduced attention to vertical guidance, while lateral deviations may reflect difficulties with navigation monitoring or trajectory control. In practice, Air Traffic Control assigns specific altitudes and routes; deviations from these clearances represent operational errors that can compromise safety.

Unlike subjective and physiological measures, performance metrics directly capture operational outcomes. However, performance degradation may occur only after cognitive overload has already developed, making these measures less sensitive to early-stage workload changes.

2.9.4 Multimodal and Composite Workload Approaches

Given the limitations of individual measurement modalities, researchers increasingly adopt multimodal approaches to assess cognitive workload and information processing (K. Li et al., 2026; Ryu & Myung, 2005). Each modality captures a different aspect of cognitive processing. Subjective measures reflect perceived effort and awareness, physiological measures capture real-time changes in cognitive demand, and performance measures indicate operational consequences. By integrating these sources, researchers can triangulate cognitive workload and reduce reliance on any single metric.

Recent studies have also explored the development of composite workload indices, which combine multiple normalized measures into a unified representation of cognitive demand (J. Yuan et al., 2025). For example, as demonstrated in recent research, physiological indicators, such as HRV features, can be standardized and combined with performance metrics, such as flight performance scores, to produce a comprehensive assessment of pilot workload (J. Yuan et al., 2025). This approach enables continuous tracking of workload while incorporating both internal cognitive state and external performance outcomes.

However, constructing composite metrics requires careful normalization and weighting of individual components. Differences in scale, variability, and sensitivity across measures must be addressed to ensure meaningful integration. Despite these challenges, multimodal and composite approaches provide a robust framework for evaluating information processing in complex operational environments

2.10 Synthesis of Literature and Research Gaps

The literature reviewed in this chapter establishes an understanding of human cognitive limitations, workload dynamics, and cockpit interface design principles in aviation. Human information-processing research demonstrates that pilots operate under strict constraints on working memory, attention, and perceptual capacity. These limitations directly affect the ability to maintain situational awareness and make accurate decisions in complex and time-critical environments.

Existing research also shows that cognitive workload emerges from the interaction between task demands and system design. Multitasking, task switching, and visual search demands increase cognitive load and degrade attention allocation. When workload exceeds available cognitive resources, performance degrades, manifesting as reduced situational awareness and an increased likelihood of operational errors.

Despite this well-established understanding of human cognitive limits, current cockpit systems continue to rely heavily on static interface architectures. Fixed display layouts and rigid information hierarchies require pilots to actively search for relevant information and integrate dispersed data elements. This design approach increases extraneous cognitive load and places additional burden on attentional resources, particularly during high-workload phases of flight.

In parallel, aviation safety systems largely follow a reactive paradigm. Procedural safeguards and alerting systems, such as TAWS, respond to hazardous states after they have already developed. These downstream interventions attempt to interrupt unsafe trajectories but do not address the underlying information processing challenges that contribute to those states. As a result, pilots must manage increasing cognitive demand while responding to late-stage warnings.

Recent advances in adaptive and context-sensitive interface design aim to address these limitations by dynamically modifying the presentation of information. These approaches demonstrate improvements in workload reduction and attention allocation under controlled conditions. However, adaptive systems often lack a systematic basis for determining which information elements to prioritize across different operational contexts. Existing approaches rarely incorporate accident-derived risk structures or data-driven models when defining information salience.

At the same time, research in data-driven safety analysis and complex network modeling provides tools for understanding how accident risk propagates through interconnected factors. Network-based approaches reveal that certain contributing factors play central roles in accident systems and can influence multiple pathways leading to unsafe states. However, these insights have not been widely integrated into cockpit interface design or information prioritization strategies. Taken together, the literature reveals a critical gap. No existing framework integrates: Established human-factors design principles and standards, Data-driven risk modeling based on accident analysis, Modular interface architecture capable of adapting to operational context, and Multimodal empirical validation using subjective, physiological, and performance-based measures.

Addressing this gap would provide a systematic basis for deciding which information should receive priority in the cockpit during a specific operational context. Instead of adding more data to the display, the proposed approach identifies the information most closely associated with accident risk and places it in stable, high-salience regions of the interface. This can reduce visual search, lower unnecessary cognitive workload, support earlier recognition of risk, and help pilots form a more accurate understanding of aircraft, terrain, and performance state before the situation becomes critical.

Previous research exists, but does not fully address this problem because the related work has developed along several paths. Human factors research explains the cognitive limits that affect pilot attention and workload. Interface design research provides principles for organizing cockpit information. Accident analysis research identifies recurring causal and contributing factors in aviation events. However, these areas are rarely integrated into a single design framework that uses accident-derived risk relationships to determine the priority of cockpit information. This dissertation addresses that gap by combining human factors design principles with data-driven risk modeling to develop modular, context-sensitive interface frameworks for aviation decision support.

3 Design Process of Generic Interface

This chapter presents the design process of a generic user interface. The chapter focuses on the placeholder layout, zones, and design rules that define where different classes of information should appear.

The generic interface architecture serves as one component of the dissertation framework. It provides a display scaffold that can support different operational contexts without changing the overall layout. Research on spatially consistent interfaces shows that stable object locations support spatial memory and improve search performance, which supports the use of a fixed interface across display states (Scarr et al., 2013).

The design process in this chapter follows a four-stage process. In stage 1, I identified source documents related to aviation displays, multifunction displays, alerting, symbology, clutter management, visual search, menu structure, and cockpit user-interface design. In stage 2, I extracted applicable requirements from those sources and organized them into a structured checklist. In stage 3, I consolidated the checklist items into a smaller set of design principles. This process produced a five-zone interface architecture consisting of a header bar, navigation bar, main content region, supplementary panel, and footer. This generic design provides the location and function of each zone but does not prescribe the operational context.

3.1 Stage 1: Source Identification

The first stage identified source documents that support the design of a generic interface. In this process, I used aviation-specific guidance as the primary design basis. The FAA Human Factors Design Guidelines for Multifunctional Display provided guidance on multifunction display layout, navigation display content, menu structure, alerting, color coding, automation, and information organization (Mejdal et al., 2001). FAA AC 25-11B supported documenting a display design philosophy, including information presentation, color use, information management, interactivity, and intended function (FAA, 2014). The same advisory circular also supported requirements for display readability, consistency, labels, symbols, and information organization (FAA, 2014).

I also reviewed MIL-STD-1472E, as it provides human engineering criteria for display systems and guidance on equipment-operator interaction (DOD, 1996). I reviewed MIL-STD-2525C because it provided guidance on symbol consistency, coding, and interpretation in a complex display environment (DOD, 2008). These standards supported the broader design goals of consistency, unambiguous coding, and rapid interpretation. Additionally, I reviewed an empirical study on visual complexity and clutter because the proposed interface is designed to reduce information-processing demands rather than add another display. The research on cockpit and tactical displays shows that clutter can increase search time and interfere with target detection, while decluttering methods such as dimming, dotting, small-sizing, and removal can improve visual search (Kim et al., 2019; Lazaro, Kang, et al., 2021; Lazaro, Kim, et al., 2021). I used these studies to support checklist items related to clutter control, salience management, and separation of primary and secondary information. Similarly, research on cockpit automation and use-interface literature related to pilot cognition has argued that cockpit interface errors often increase cognitive demand when pilots must recall memorized action sequences rather than rely on visual labels, prompts, and feedback

(Sherry et al., 2002). Table 1 summarizes the source categories used in the design process.

Table 1 Source documents used to develop the generic interface

Source Category	**Source Reviewed**	**Design role**
FAA aviation display guidance	1) FAA Human Factors Design Guidelines for Multi-function displays 2) FAA AC 25-11B	Supported requirements for display layout, philosophy, information management, intended function, alerting, labels, consistency, color coding, and information organization.
Military human-engineering standards	MIL-STD-1472E	Human-engineering criteria for displays, controls, labels, layout, coding, and operator-equipment interaction.
Military and tactical symbology standards	MIL-STD-2525C	Symbol consistency, tactical display coding, and standardized visual representation of operational information
MFD organization literature	(Seidler & Wickens, 1992)	Grouping related features, layout simplification, fast and clear navigation, and clear menu hierarchies
Cockpit automation and UI literature	(Sherry et al., 2002)	Visible labels, prompts, and feedback
Visual complexity and decluttering literature	(Kim et al., 2019; Lazaro, Kang, et al., 2021; Lazaro, Kim, et al., 2021)	Clutter control, dimming, small-sizing, task-based information prioritization
Spatial consistency literature	(Scarr et al., 2013)	Fixed zone layout and the need to preserve stable object locations across interface states
Display form literature	(Shupp et al., 2009)	Display size, curvature and accessibility of distributed information

Quantitative clutter	(Q. Liu et al., 2023)	Information density reduction
General user-interface and human-factors design guidance	General usability and accessibility recommendation	Visual grouping, flexibility, feedback

3.2 Stage 2: Checklist Development

The second stage converted the sourced documents mentioned in Table 1 into a structured design checklist. For each source, I extracted guidelines that reduced working memory demand and cognitive load, improved information processing and visual search, enabled faster alert detection, reduced display clutter, improved situation awareness, and reduced spatial inconsistency. After requirements are extracted, each item is categorized by priority and applicability. This stage excluded guidance on display functions outside the scope of this dissertation's objective, such as hardware certification, ATC displays, or weather displays. The checklist used three priority levels: critical, recommended, and optional. The guidance that directly affected safety and workload reduction and was identified as mandatory is classified as critical. Similarly, if the guidance improves workload management, supports human factors practices, and reduces cognitive strain, it is classified as recommended. Context-dependent enhancements that support flexibility are classified as optional. The checklist also incorporated findings from visual-search and decluttering studies, such as hiding information that may become relevant later and small-sizing or dimming low-priority information.

Table 2 Checklist used to derive the generic interface architecture

Category	Extracted guidance	Priority	Design Implication
Alerting Signal	Position visual alerts close to pilot's line of sight	Critical	Reserve a high salience region for critical alerts and notifications
	Prioritize alerts by urgency levels		Separate warning, caution, and advisory logic through position, color, and wording
	Use conspicuous visual features for high-priority alerts		Separate region for information that requires immediate attention.
	Minimize low-priority advisory alerts	Recommended	Keep advisory information away from the most salient alert region
Menu Structure	Provide main menu and navigation functions	Critical	Include a stable navigation region that supports movement across interface functions.
	Limit menu breadth to prevent overload		Avoid crowded menus and use a small number of stable navigation options
Menu organization	Group menu items into meaningful categories	Critical	Navigation items in a dedicated region, avoid mixing with primary operational information
Menu labels	Avoid generic categories such as “miscellaneous”	Critical	Use explicit labels that support recognition rather than recall
Layout and arrangement	Group functionally related components together	Critical	Divide the interface into functional zones
	Display only required information to avoid clutter		Display of continuous primary information in the main content and secondary information elsewhere
	Place the most important components in the most visible locations.		Alerts and system status information should be in the user's line of sight, mainly centered on the header
	Separate distinct categories of data from each other		Use boundaries between different functional groupings

Visual hierarchy	Use consistent coding to indicate information importance	Critical	Use position, color, and size and grouping to guide attention
Decluttering	Reduce salience of lower-priority information without hiding information needed for later review	Recommended	A secondary visual treatment for supporting information
Spatial consistency	Keep major display regions stable across interface states	Critical	Preserve a fixed layout
Information Presentation	Collocate items that the human must mentally integrate	Recommended	Place-related information that the user can integrate into the same operational regions
Feedback and Status	Keep users informed about the system performance and the results of system actions	Recommended	Include low-priority system state information in stable regions

The checklist converted broad and various guidelines into concrete interface design constraints. I based this classification on normative language within standards (e.g., “shall”, “should”, “may”) and on the documented cognitive impact of each requirement. I removed duplicates, resolved overlaps across documents, and consolidated recurring themes into a unified set of architectural constraints. For example, multiple sources independently emphasize clutter control, grouping related elements, and layout stability. I synthesized the structural principles governing spatial zoning.

3.3 Stage 3: Translation of Requirements to Spatial Architecture

I derived the spatial interface structure directly from the consolidated checklist. The checklist revealed recurring structural constraints governing visual salience, grouping, information density, and alert prioritization. I translated these constraints into a fixed zoning architecture that regulates where information categories may appear.

Five dominant architectural requirements emerged during consolidation. First, time critical alerts require a stable, high-salience location that supports rapid visual detection. Second, frequently sampled and safety-critical information must occupy stable and predictable locations to minimize visual search cost. Third, functionally related elements must remain collocated to reduce cognitive integration effort. Fourth, contextual or lower-priority information must remain accessible without increasing baseline information density. Fifth, low-priority system-state information must remain visible without competing with primary operational information. These architectural requirements translated into interface zones as shown in Table 3.

Table 3 Translation of design principle to interface zones

Design Principle	**Extracted requirements**	**Translation**	**Interface zones**
Alert prioritization guidance	Time-critical alerts require high salience and persistent visibility	Dedicated high-salience alert area	Header bar
Visual search and attention principles	Frequently sampled information should remain in stable location	Stable primary operational region	Main content region
Grouping and proximity principles	Functionally related information should remain collocated	Related operational variable grouped together	Main content region and secondary information panel
Consistency and spatial mapping	The layout should remain predictable across task and context changes	Fixed zoning architecture	All zones
Configurability principles	User controls should remain accessible but should not interfere with task-critical cues	Dedicated configuration and interface navigation regions	Navigation sidebar

Status-awareness principles	Low priority system-state information should remain visible but non-intrusive	Dedicated low-salience status regions	A low priority (footer) regions

I mapped these requirements into a five-zone spatial architecture, as shown in Figure 5. Header Bar- Dedicated to critical alerts and global system status. This placement satisfies central field-of-view and salience constraints and ensures persistent visibility of system state. Main Content Region- Reserved for primary operational information and dynamically prioritized mission-specific content. This region supports frequent sampling and direct attention allocation. Navigation Sidebar- Provides stable access to configuration controls, filters, and workflow navigation while preserving layout invariance across contexts. Supplementary Panel- Contains contextual details, expanded data views, and lower-priority information that should remain accessible without competing with primary task data. Footer: This region contains low-priority information, such as system time, version information, or data source status. This zoning structure enforces visual hierarchy while limiting cross-competition between information categories. It preserves spatial consistency across operational contexts, reducing cognitive reorientation costs when task demands change. The architecture remains fixed. Only the content populating each zone varies depending on the mission context.

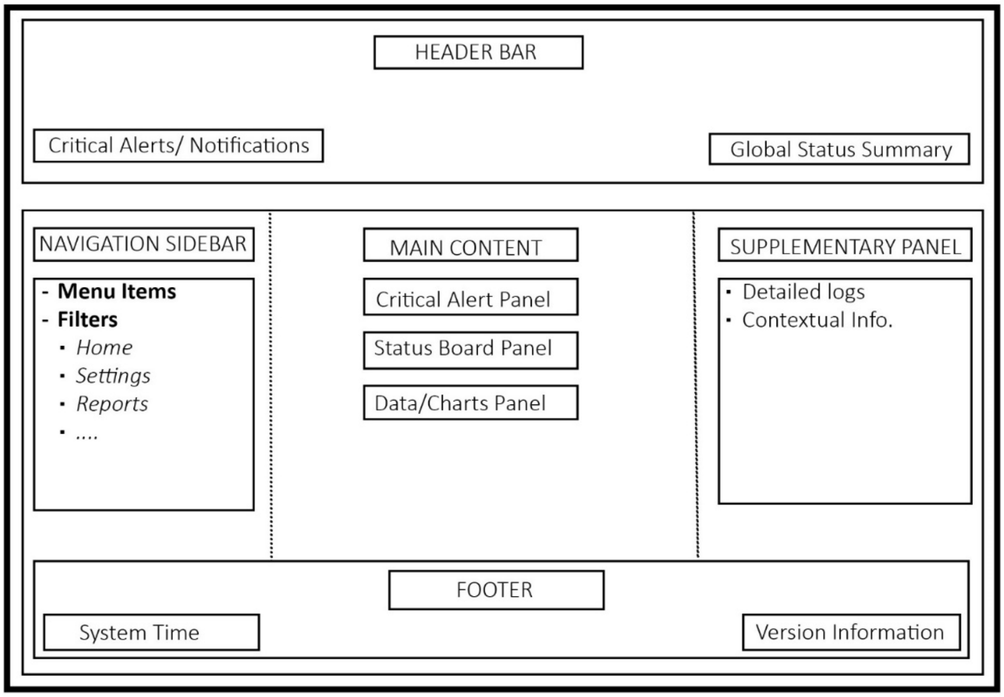

Figure 5 Generic interface architecture with fixed zones for context-specific information

Table 4 summarizes each interface zone, its purpose, cognitive rationale, and the content allowed and prohibited

Table 4 Different types of information that are allowed and prohibited in the interface

Interface Zone	Purpose	Cognitive Rationale	Allowed Content	Prohibited Content	Update Behavior
Header Bar	Presents a critical alert and global system status	Keep urgent information in a stable, high-salience location	Warning, Caution, Advisor, global status	Detailed raw data, secondary menus, and low-priority information	Event-driven and continuously visible
Main Content Region	Presents primary mission-critical operational information	Reduce visual search and support situation awareness	Primary operational display, or mission visualization, core decision support metrics	Configuration menus, secondary tables, non-critical status	Context-sensitive population based on dominant operational risk
Navigation Sidebar	Presents stable access to control and workflow navigation	Preserves layout consistency and reduces reorientation cost	Filters, display modes, configuration controls	Time-critical alerts or primary flight-risk information	User-initiated or mode selection update
Supplementary Panel	Presents contextual or expanded information	Keep lower-priority accessible without cluttering the primary display	Expanded table, explanation, secondary cues	Immediate hazard or primary control cues.	Context-sensitive and expandable.

Footer/status regions	Presents low-priority system-state information	Keeps system-state information visible without competing with primary operational information	System time, software or data version, low priority status	Critical alerts, primary operational values, dense tables, configuration menus	Continuously visible

3.4 Separation of Architecture and Mission-Specific Content

The generic interface separates spatial structure from mission-specific information content. The header bar, main content region, navigation sidebar, supplementary panel, and footer/status region remain structurally invariant across operational contexts. This invariance preserves spatial predictability and reduces cognitive reorientation cost during phase transition or mission changes.

The architecture defines where categories of information may appear, but it does not prescribe what specific data must populate each zone. Mission-specific variables were later populated into the predefined regions based on operational context and prioritization logic, while the structural layout remained unchanged.

For example, operations associated with CFIT risk, terrain proximity metrics, and climb performance cues populate the primary operational region. The header bar displays alerts within the pilot's central field of view. Contextual information, such as expanded performance tables, appears in the supplementary panel to remain accessible without competing for primary attention. Similarly, if the operational context shifts, for example, from mountainous terrain to flat terrain en route during cruise, the terrain margin no longer occupies the central region. Instead, fuel endurance margin, weather penetration risk, or traffic conflict geometry (in high-density terminal airspace) are all in the same primary zone. The header continues to display high-priority alerts, but the alert content reflects the dominant variables of the current context rather than terrain proximity.

The same architecture can support a different operational context if the designer performs a new risk-identification process. For example, in an en route cruise context over non-mountainous terrain, terrain margin may no longer require central placement. A different analysis may identify fuel endurance margin, weather penetration risk, traffic conflict geometry, or system-degradation status as the dominant operational variables. Those variables could then populate the same predefined zones. The header would still display high-priority alerts, the main content region would still display primary operational information, and the supplementary panel would still provide contextual detail.

The generic interface permits controlled adaptation within predefined structural constraints. Zones do not relocate, expand unpredictably, or reorganize across contexts. The generic interface is a fixed-layout, context-sensitive, rule-based, adaptive interface. The term fixed layout indicates that the spatial zones remain invariant across operational contexts. Context-sensitive refers to substituting mission-relevant information within those predefined zones based on the operational environment. Adaptive is used in a bounded sense: the interface updates content, thresholds, or risk prioritization, but it does not dynamically reorganize the spatial layout. Therefore, the framework adapts the information content without altering the display's physical architecture.

By maintaining structural invariance and allowing mission-sensitive content substitution, the framework achieves domain adaptability without introducing layout instability or additional cognitive burden. The generic interface provides the structural foundation for mission-specific prioritization mechanisms. In subsequent chapters, I apply complex network analysis of accident data to identify the dominant risk factors within the predefined zones.

The generic interface is considered transferable to another operational domain only if the spatial architecture remains unchanged while the mission-specific content is replaced through a new risk-identification process. When moving from mountainous CFIT operations to another domain, the designer must redefine the operational context and construct or identify the relevant risk factors, then prioritize those factors using an appropriate analytical method, and map the resulting high-priority variables into the existing zones. The elements that may change are the risk dataset, network nodes, priority variables, thresholds, and displayed content. The elements that must remain fixed are the zone structure, salience hierarchy, grouping logic, and rules governing information density. If a new domain requires relocating the zones themselves, then the framework has not generalized as intended.

3.5 Distinction from Existing Cockpit Architectures

Modern glass cockpits comply with regulatory display standards and provide extensive flight, navigation, terrain, and systems information. However, these systems typically follow fixed certification-driven layouts in which information categories are statically allocated to display regions. While alerting systems such as TAWS and Traffic Collision Avoidance System (TCAS) introduce reactive warnings when thresholds are violated, the baseline spatial organization does not dynamically prioritize information based on mission-specific risk patterns.

In conventional cockpit configurations, terrain information, traffic displays, engine parameters, and navigation data coexist statically within predefined instrument groupings. The relative salience of these elements often depends on reactive alert triggering rather than on contextual risk weighting. As a result, safety-critical variables may compete visually with secondary information until an exceedance condition occurs.

The proposed generic interface differs in three fundamental ways:

1. Risk-Driven Prioritization Before Threshold Violation

   The architecture supports the proactive elevation of dominant risk variables into high-salience zones based on operational context. For example, during mountainous approach operations, altitude margin and terrain proximity metrics occupy the primary operational region before a warning threshold is crossed.

2. Separation of Architecture and Content

   Conventional cockpits rigidly couple layout and information categories. The proposed framework decouples structural zoning from mission-specific content. The spatial layout remains invariant, while the dominant risk variables adapt to context.

3. Bounded flexibility within stable spatial memory

   The design preserves spatial consistency while allowing controlled substitution of risk-relevant metrics. This avoids layout instability while preventing the prioritization of static information.

The distinction is not graphical or aesthetic. The difference lies in how information priority is determined and spatially expressed. Existing systems primarily rely on reactive alerts layered atop fixed layouts. The generic interface introduces a structured method for proactive, context-sensitive adaptation within a stable architecture scaffold.

3.6 Summary

This chapter presented a design process used to develop a generic interface. I derived the spatial architecture from consolidated human-factors standards, interaction design principles, and workload management constraints. I translated these requirements into a fixed zoning structure that preserves spatial invariance while enabling mission-sensitive adaptation.

The generic interface defines *where* categories of information may appear and establishes architectural constraints that govern salience, grouping, and density. It does not determine which specific risk variables populate those regions. The identification and prioritization of mission-critical variables require a separate analytical process.

Chapter 4 introduces the process. I analyze CFIT accidents using complex network theory to identify dominant structural risk factors. I then map these risk drivers to the predefined architectural zones established in this chapter. This separation ensures that architectural design and risk identification remain analytically distinct while operationally integrated.

4 Controlled Flight into Terrain

As established in Chapter 3, the design framework provides a domain-agnostic method for creating a context-sensitive interface. To demonstrate its practical application, this research applies the framework to a particularly challenging and high-stakes aviation scenario, CFIT in mountainous regions.

This chapter first justifies the selection of CFIT as the case study by highlighting its severity, its direct relevance to information-processing errors, and the clear technological gap in current cockpit systems. I then detail how I executed the data-driven analysis to populate the generic interface. The outcome of this process is a tailored interface that I designed to enhance pilot decision-making and prevent CFIT accidents.

4.1 CFIT Overview

CFIT is a unique aviation occurrence in which a mechanically fit aircraft, under full pilot control, accidentally collides with terrain, water, or obstacles despite the aircraft’s operational capabilities and the crew's ability to maneuver. International Civil Aviation Organization (ICAO) defines CFIT as “A situation where a properly functioning aircraft under the control of a fully qualified and certificated crew is flown into terrain (mountain, ground, water mass, trees, etc.) with no apparent awareness on the part of the crew” (FAA, 2003). CFIT accidents represented 5.35% of the total accidents from 2005 to 2023 (IATA, n.d.). CFIT consistently exhibits a higher fatality risk than other accident categories (e.g., LOC-I, Runway Excursion), although its accident rate is relatively low. However, when CFIT accidents occur, they are catastrophic, which increases the fatality risk relative to other accident categories, such as LOC-I, Runway Incursions & Excursions, and System Component Failures (SCF).

These occurrences are predominantly attributed to errors arising from nonadherence to or failure to cross-verify SOP (Baker & Lamb, 1989; FAA, 2003). An analysis of Quick Access Recorder (QAR) overrun triggers from Sichuan Airlines flight operations in China revealed that CFIT occurrences increase when crew members violate established rules, including flying below the minimum safe altitude, failing to take corrective action after a warning, incorrectly executing ATC instructions, and deliberately ignoring SOPs (Tian et al., 2022).

Air Traffic Services and adverse meteorological conditions are two of the most common threats for CFIT. Tian et al. (2022) identified low visibility and Instrument Meteorological Conditions (IMC) as major meteorological contributors for CFIT. IMC refers to weather where visibility is too poor for pilots to see the ground or a natural horizon, legally requiring them to fly by instrument alone. The danger arises when pilots accustomed to flying under Visual Flight Rules (VFR) encounter these conditions, as the loss of all visual references can lead to spatial disorientation and loss of aircraft control (Damiba, 2024; Lee, 2012). International Air Transport Association (IATA) provides a breakdown of air traffic services threats, such as challenging clearances, reroutes, language barriers, controller errors, and separation failures, though they do not specify the individual frequency or prevalence of each factor (IATA, n.d.).

CFIT remains a significant safety challenge in aviation, particularly because it poses a high risk of fatalities despite a relatively low accident rate. The combination of crew errors, lapses in adherence to SOPs, adverse meteorological conditions, and challenges in air traffic services has underscored the complexity of CFIT occurrences. While various safety initiatives have reduced the overall frequency of such events, the continued occurrence of CFIT events reveals a persistent vulnerability in aviation safety. This vulnerability becomes even more pronounced in mountainous operations. NTSB data from 2008 to 2022 highlights the risk. As shown in Figure 6, mountain and hilly terrain was the second most common location for terrain-related occurrences (311), surpassed only by rough, uneven terrain (392). These occurrences reflect the amplified risk in such a challenging environment. The following section explores key factors that amplify risk in challenging environments. It will examine the roles of complex terrain, dynamic weather conditions, and human performance limitations, and analyze how decision aids can play a transformative role in mitigating risk.

Altitude further compounds these challenges. Human cognitive performance decreases with increasing altitude, particularly in unpressurized aircraft (K. Martin et al., 2019). Hypoxia impairs attention, working memory, and decision-making. Although individual susceptibility varies based on age, smoking history, and recent altitude exposure, overall cognitive performance declines with increasing altitude (K. Martin et al., 2019). Despite the availability of technologies such as TAWS, which provide terrain proximity alerts to prevent ground collisions, CFIT in mountainous regions remains highly lethal, particularly in general aviation.

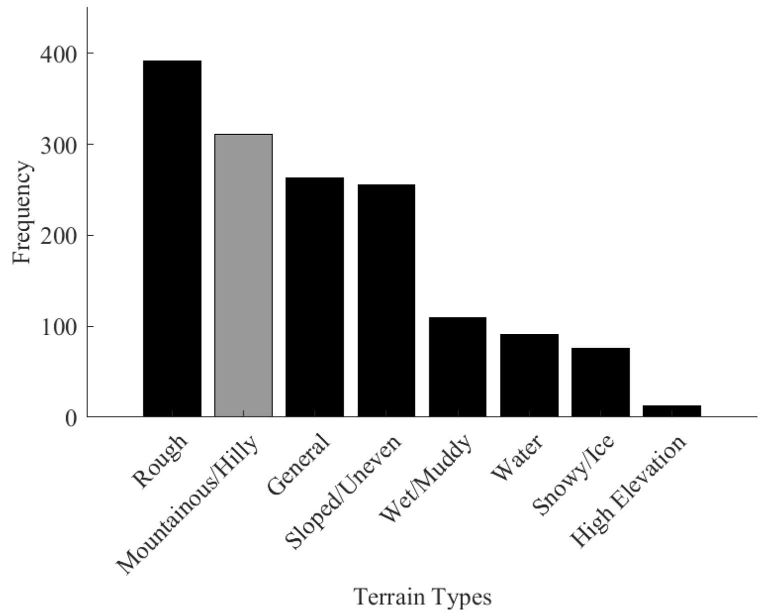

Figure 6 Bar graph showing mountainous hilly regions as one of the most common causes for CFIT occurrences

4.2 CFIT in Mountainous Regions

The complex topography of mountainous regions, characterized by steep gradients, winding paths, and frequent elevation changes, significantly reduces pilots' visibility and situational awareness. Mountainous terrain disrupts all three levels by obscuring visual cues, increasing environmental variability, and reducing the time available for accurate projection.

This terrain generates mesoscale weather systems that trigger highly localized extreme conditions, such as trapped mountain waves, low-level cloud ceilings, supercooled cloud water, hail, strong tailwind gradients, and turbulence (Regmi et al., 2017). Airports in these regions are often located in deep valleys, narrow river basins, or short mountain strips. Air routes connecting these airports frequently cross high ridges and narrow gaps, then descend through winding river gorges (Regmi et al., 2017).

Rapidly changing weather and terrain-induced constraints require continuous monitoring, precise navigation, and rapid decision-making.

4.3 Cognitive Performance and Decision-Making in Mountainous Regions

Mountain flying requires high cognitive demands due to complex terrain, rapidly changing mesoscale weather systems, and altitude-related physiological stressors. These factors increase pilot workload and directly impair decision-making performance. In mountainous environments, pilots must maintain high navigation accuracy while continuously adjusting attitude and heading to clear steep slopes, narrow gorges, and evolving weather systems, including cloud formation and precipitation (Troller et al., 2016). Frequent heading and attitude corrections across multiple flight segments further increase workload (Davis et al., 2005).

As established in Chapter 2, pilots rely on limited working memory and attentional resources to perceive, interpret, and project flight conditions. When task demands exceed this cognitive capacity, working memory overload occurs. Under high-workload conditions, pilots struggle to process concurrent information streams, increasing the likelihood of attentional tunneling, cue misinterpretation, and delayed response execution.

In mountainous operations, this overload can manifest as denial of emerging threats, cognitive freezing, or exaggerated startle responses (Ladurini et al., 2024; W. L. Martin et al., 2011). These cognitive disruptions delay critical decisions and degrade maneuver execution, thereby increasing the probability of CFIT.

These cognitive limitations do not occur in isolation. Interface design directly influences workload, attentional allocation, and decision-making performance (Curtis et al., 2010; Wickens & Carswell, 2021). When an interface fails to prioritize terrain and performance information under high workload, pilots must compensate by reallocating limited attentional and working memory resources, which increases intrinsic cognitive load and reduces information processing efficiency (Sweller, 1988; Wickens & Carswell, 2021).

4.4 Technological Gaps

Technological advancements have reduced the overall frequency of CFIT accidents. Systems such as TAWS provide terrain proximity alerts intended to prevent ground collisions. TAWS integrates altitude, vertical speed, and terrain database information to generate visual and auditory warnings when the aircraft approaches hazardous terrain (Lee, 2012).

Despite these improvements, CFIT accidents continue to occur, particularly during mountainous operations and non-precision approaches. In non-precision approaches, pilots do not receive vertical guidance and must rely heavily on situational awareness, manual altitude management, and judgement (Breen, 1999; Bud et al., 1997; Khatwa & Roelen, 1999; Lee, 2012).

TAWS remains fundamentally reactive. It issues an alert only after the aircraft approaches the predefined terrain threat threshold. It does not proactively restructure the information environment to reduce cognitive load before the risk materializes. In high-workload environments such as mountainous terrain, reactive alerts may compete with other task demands, increasing cognitive strain rather than reducing it. Moreover, installing and upgrading advanced systems such as TAWS and Enhanced Ground Proximity Warning Systems (EGPWS) imposes a significant financial burden (Lee, 2012). Many general aviation operators continue to use older systems that provide limited predictive capability and generate nuisance alerts (Bud et al., 1997; Honeywell, 2000). For example, in an accident that occurred on Mustang, Nepal, involving a DHC-6/300 on 29th May 2022, the pilot had effectively disabled or misconfigured the aircraft's terrain warning system because the crew found its alerts distracting. As a result, the system did not issue a real warning in time to prevent a crash, and the pilot failed to take the corrective action needed to avoid disaster (MoCTCA, n.d.).

Current systems, therefore, address terrain proximity but do not systematically prioritize mission-critical information based on historical accident patterns. They warn pilots of imminent danger, but they do not proactively guide decision-making during high cognitive demand. This gap motivates the development of a data-driven interface framework that reorganizes the presentation of information before workload escalates to unsafe levels.

4.5 Data-Driven approach in identifying critical factors

A single, linear cause does not drive CFIT accidents; pilot behavior, aircraft performance, and environmental hazards interact in ways that can form circular feedback loops (D. Kelly & Efthymiou, 2019; Meng & Lu, 2022). For example, poor situational awareness due to poor visibility (an environmental hazard) can lead a pilot to select incorrect control inputs, resulting in an initial piloting error. Incorrect input may cause the aircraft to descend or turn, bringing it closer to the terrain. This dangerous state increases pilots' stress and workload, thereby degrading situational awareness and increasing the risk of additional errors. This self-reinforcing cycle is what can lead to failure. Traditional linear models, such as Failure Mode and Effects Analysis (FMEA) and the 2-4 model, and systems-based approaches, such as STAMP, capture important elements (Ericson, 2005; Fu et al., 2013; Leveson, 2004). However, they may not fully describe these tightly coupled interactions in a way that reveals critical propagation pathways. Complex network theory treats risks as interconnected nodes and edges, and has successfully revealed causal structures in many safety domains (Lin et al., 2014; P. Wang et al., 2018). Researchers have integrated complex network theory, grounded theory, and Bayesian networks into accident studies. They first apply grounded theory to systematically extract and code causal factors from narrative accident reports. Later, they construct network models to represent structural relationships among these factors and use Bayesian networks (Yue et al., 2021). Other studies have integrated complex network models with Bayesian networks to estimate probabilistic dependencies and simulate risk propagation among contributing factors (Zhang & Mahadevan, 2021).

Therefore, this study proposes a method that applies complex network theory to NTSB data from 2008 to 2022 to identify critical nodes and their relationship in CFIT mountain accidents. By mapping this relationship, I aim to reveal high-leverage points for intervention and lay the groundwork for data-driven safety enhancements in mountain environments.

4.5.1 Methodology

I preprocessed the NTSB accident dataset by constructing a co-occurrence frequency matrix from the codes assigned to each Event ID. Accidents where mountain/hilly terrain appeared as a causal or contributing factor were included in the analysis. However, non-CFIT cases, incomplete records, and duplicate Event IDs were excluded. Each accident report includes a unique Event ID and one or more finding codes that represent causal or contributing factors, categorized as personnel, aircraft, or environmental issues. When multiple finding codes appeared under the same Event ID, I treated them as co-occurring factors.

I then computed the frequency with which each finding code co-occurred with every other finding code across all accidents. This process produced a symmetric frequency matrix in which rows and columns represent individual finding codes. The diagonal elements represent the total occurrence frequency of each finding code, and the off-diagonal elements represent the number of times two finding codes appeared together within the same accident.

Table 5 An Event ID can contain multiple finding codes

Event ID	**Finding codes assigned to the event**
20201005102094	0206000044 0302101091 0206110104 0303102091
20210628103362	0302101091 0302202582

These event-level finding codes produce the co-occurrence frequency matrix shown in Table 6. The analysis excluded diagonal entries because they represent the individual frequency of occurrence of a finding code rather than a relation between two finding codes.

Table 6 The co-occurrence frequency matrix is created from finding codes

Finding Code	**0206000044**	**0302101091**	**0206110104**	**0303102091**	**0302202582**

0206000044	1	1	1	1	0
0302101091	1	2	1	1	1
0206110104	1	1	1	1	0
0303102091	1	1	1	1	0
0302202582	0	1	0	0	1

I defined each unique finding code as a node in the network. I then constructed an undirected edge list from the off-diagonal entries of the co-occurrence frequency matrix. Each edge connects two distinct finding codes that appeared together within at least one accident Event ID. To account for differences in overall accident counts, I normalized each edge weight by dividing the co-occurrence count between two finding codes by the total number of observed co-occurrence instances. This normalization produces a weighted, undirected network in which edge weights range from 0 to 1. Higher co-occurrence frequency indicates stronger weighted connections among contributing factors.

The NTSB maintains a publicly accessible accident database containing structured information on aircraft, environmental conditions, personnel, findings, probable causes, and contributing factors (NTSB, n.d.). I used the structured finding codes in this database and employed an interpretation approach similar to that described by Fala (2022). The NTSB database uses a unique 10-digit finding code for each finding for accidents after 2008. The system breaks this code down into 5 distinct groups: *category, sub-category, section, subsection, and modifier codes,* which together classify and characterize an occurrence.

Table 7 provides an example of how one can interpret this coding schema using the finding code 0302101091. The code's structure breaks down as follows. The first two digits (03xxxxxxxx) identify the main category, in this case, "Environmental Issues". The next two digits (xx02xxxxxx) specify the subcategory, "Physical Condition". The following two pairs of digits define the section as "Terrain" (xxxx10xxxx) and the sub-section as "Mountain/Hilly" (xxxxxx10xx). The final two digits (xxxxxxxx91) are the modifier, which characterizes the finding's role in the accident. Here, the modifier means the mountain/hilly terrain "contributed to the outcome." I filtered events with category number 03 (environmental issue), subcategory number 02, section number 10 (terrain), and subsection number 10 (Mountain/Hilly) for this analysis, as this combination encompasses all incidents where mountain/hilly terrain was a factor. This process yielded 287 events used in this study.

Table 7 Description of Finding Code "0302101091"

Finding Code	**Category (category_no)**	**Subcategory (subcategory_no)**	**Section (section_no)**	**Subsection (subsection_no)**	**Modifier**
0302101091	03	02	10	10	91
	Environmental Issue	Physical Condition	Terrain	Mountain/Hilly	Contributed to Ouctome

4.5.1.1 Complex Network Theory Metrics

In this study, I modeled accidents as a complex network. Complex network theory is appropriate for this study because the NTSB dataset contains structured co-occurrence relationships among multiple contributing factors within each accident. These factors do not operate independently; they appear simultaneously and interact within the same event. This multi-factor interaction structure aligns naturally with network modeling, in which nodes represent discrete factors and edges represent observed co-occurrence.

Unlike linear causal chains, this representation allows me to quantify how often factors cluster and identify highly interconnected nodes that may serve as structural hubs in accident propagation. Therefore, CNT provides a mathematically consistent framework for modeling the interconnected structure inherent in CFIT accident data. Within this network, I treat individual causal or contributing factors (e.g., pilot decisions or weather hazards) as nodes and represent documented co-occurrences between them as edges. Metrics characterize the network's overall structure and dynamics. These measures provide quantitative metrics for comparing the importance of different elements within the network. Specifically, these metrics quantify network structural properties, identify nodes with high degree centrality (hubs), and detect nodes with high betweenness centrality that serve as bridges between clusters of contributing factors.

The diameter is the longest shortest path between any two nodes in the network. i.e., the minimum number of steps it takes to connect the two most distant nodes within the network (K. Li & Wang, 2018; J.-F. Yang et al., 2023; Yue et al., 2021). A smaller diameter indicates a less complex network, as all factors are relatively close to one another.

Graph density is the ratio of the actual number of existing connected edges to the maximum possible number of connected edges in a network. Graphs with a density closer to one are considered dense, while those with a value closer to zero are considered sparse. Density can be calculated using Equation 1 (S. Li et al., 2025)

$$D = \frac{2E}{N(N-1)} \quad \text{Equation 1}$$

where E represents the actual number of connected edges in the network, and N represents the number of nodes in the network.

The average path length measures the communication efficiency and closeness of the entire network. A lower average path length indicates a more efficient network, meaning information and influence can pass between factors through fewer nodes (S. Li et al., 2025). I calculate this value by measuring the shortest path between all pairs of nodes in the network and then averaging them, as shown in Equation 2

$$L_s = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij} \quad \text{Equation 2}$$

where d_{ij} represents the shortest path length from node i to j

Betweenness centrality measures the extent to which a node lies on the shortest paths between any two other nodes. This metric is crucial for understanding how information and influence flow within the network. Nodes with high betweenness centrality act as important bridges, and, therefore, play a critical role in the spread of risk through the network (J.-F. Yang et al., 2023). I calculate betweenness centrality using Equation 3

$$b_m = \sum_{i,j \epsilon N, i \neq j} \frac{n_{ij(m)}}{n_{ij}} \quad \text{Equation 3}$$

where b_m represents the betweenness centrality, which passes through the node m to the total number of shortest paths among all shortest paths, n_{ij} represents the number of all shortest path connections between node i and node j in the network.

4.5.2 Network Results

Figure 7 shows the network model depicting the interconnection between causal/contributing factors for CFIT-Mountainous Terrain accidents. The network includes 282 nodes and 1223 edges, and has a network diameter of 5, an average path length of 2.39, and a network density of 0.031. A density of 0.031 means that only about 3% of all possible connections between contributing factors were present in the accident database. In other words, most contributing factors do not directly appear with most other factors in the same accident reports. Density indicates that the network is relatively sparse, meaning that only specific combinations of factors tend to occur together rather than all factors being widely interconnected.

The average path length of 2.39 indicates that most pairs of factors are only two to three steps apart in the network. Even if two factors do not appear together directly, they are usually connected through a small number of shared associations. In practical terms, this shows that CFIT accidents in mountainous terrain rarely involve a single isolated issue. Instead, a small cluster of related problems tends to occur within the same event. For example, degraded information processing, altitude mismanagement, and aircraft control difficulties often co-occur rather than occur independently.

In Figure 7, node size represents how frequently each finding code appears in the dataset. Larger nodes indicate factors that occur more frequently in CFIT accidents in mountainous terrain. Edge thickness represents the normalized co-occurrence strength between two factors. Thicker edges indicate that two contributing factors co-occur more frequently in the same accidents.

Nodes A1 (Altitude not maintained), B1 (Decision-making/Judgment by pilot), and B2 (Aircraft control) appear larger because they co-occur with many other contributing factors. The thicker edges among these nodes indicate that they frequently co-occur in the same accidents relative to other factor combinations. This pattern suggests that these factors play a central role in CFIT operations in mountainous terrain.

Three centrality measures are used because each captures a different form of structural importance within the CFIT accident network. Degree centrality identifies the finding codes that co-occur with many other factors and therefore represent broadly connected accident contributors. Weighted degree centrality extends this logic by incorporating the frequency of co-occurrence, allowing the analysis to distinguish between factors that are merely connected to many nodes and those that repeatedly appear in common accident patterns. Betweenness centrality identifies factors that lie on the shortest paths between other nodes and therefore serve as structural bridges between different clusters of accident contributors. Together, these measures support the identification of factors that are broadly connected, frequently recurring, and structurally positioned to link different accident mechanisms.

However, these centrality metrics do not establish direct causality or temporal ordering among accident factors. Because the network is undirected and based on co-occurrence within accident reports, a high-centrality node should be interpreted as structurally prominent rather than causally dominant. Therefore, the results are used to prioritize interface content and safety attention, not to claim that one factor directly caused another.

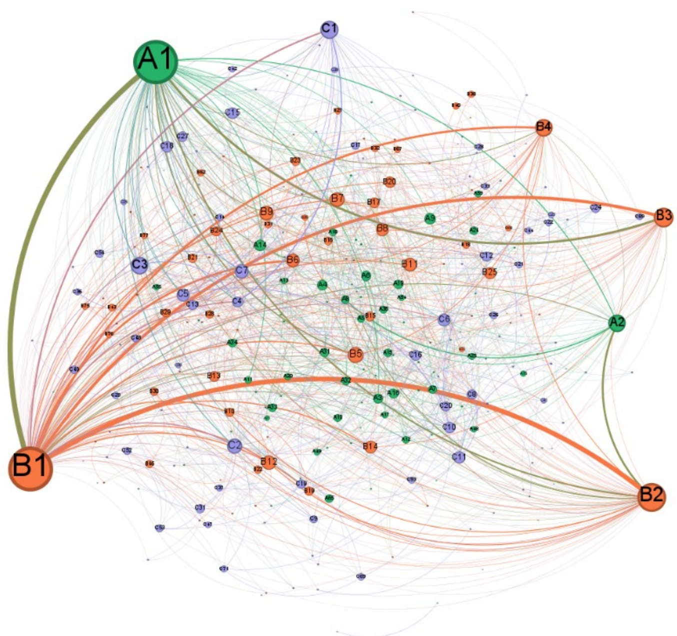

Figure 7 Network representation of CFIT-Mountainous/Hilly Operations. Nodes represent contributing factors, and edges represent their co-occurrence within accident reports. Orange nodes represent personnel issues, purple nodes are environmental issues, and green nodes represent aircraft issues

4.5.2.1 Degree Centrality

The degree of a node is the number of direct connections (or edges) it has to other nodes (Cherven, 2015). Nodes with higher degree centrality co-occur with many other contributing factors and therefore occupy structurally central positions within the network. In this study, I calculated degree centrality to quantify the number of other factors to which each finding code directly connects in the co-occurrence matrix. The node with the highest degree centrality was B1 (Personnel issues-Action/decision-Info processing/decision-Decision making/judgment), A1 (Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Altitude), B2 (Personnel issues-

Task performance-Use of equip/info-Aircraft control), B3 (Personnel issues-Psychological-Attention/monitoring-Monitoring environment), A2, C6, B4, B5, C8, and C7, with degree values of 143, 129, 84, 46, 38, 36, 35, 34, 32, and 29, respectively, as shown in Figure 8. Appendix A provides a detailed definition of a node assigned to a finding code.

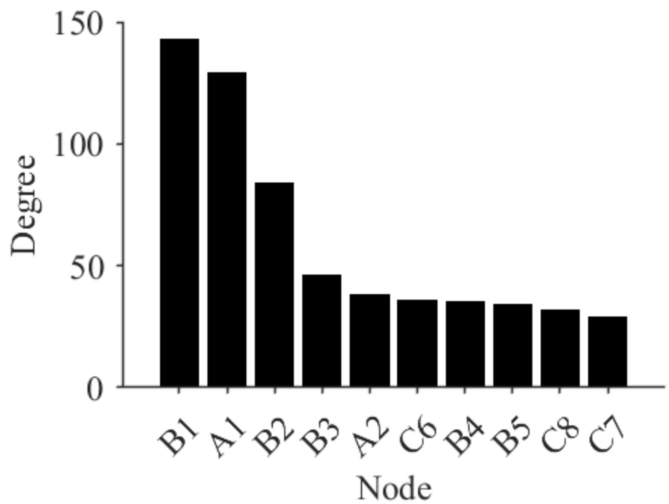

Figure 8 Degree centrality of the ten highest ranked nodes

These high-degree nodes frequently co-occur with multiple other contributing factors in the same accident reports. Their central position in the network indicates that CFIT mountainous accidents consistently involve these factors alongside other issues rather than in isolation. In particular, B1, A1, and B2 form a tightly interconnected group, suggesting that cognitive and control-related factors commonly cluster within these accidents.

Figure 9 shows the cumulative degree distribution of the network, which follows a power-law trend. The relationship between the cumulative degree Y and node degree X follows a function $\log_{10} Y = 1.90 \log_{10} X - 1.466$. The approximately linear relationship in log-log space is consistent with a heavy-tailed degree distribution, in which most nodes have few connections while a small number of nodes are highly connected. I describe the distribution as heavy-tailed rather than strictly scale-free, because a linear fit in log-log space does not, by itself, constitute a formal test of the power law hypothesis (Clauset et al., 2009). A definitive classification would require maximum-likelihood estimation of the scaling exponent and a goodness-of-fit comparison with alternative heavy-tailed distributions, such as the log-normal or exponential. For the purposes of this study, the relevant structural observation is that co-occurrence is concentrated among a small number of dominant nodes rather than evenly distributed across factors, which is sufficient to justify prioritizing those nodes in the interface.

In practical terms, this means that CFIT accidents in mountainous terrain do not distribute risk evenly across all contributing factors. Instead, a small number of highly connected factors dominate the network's structural organization.

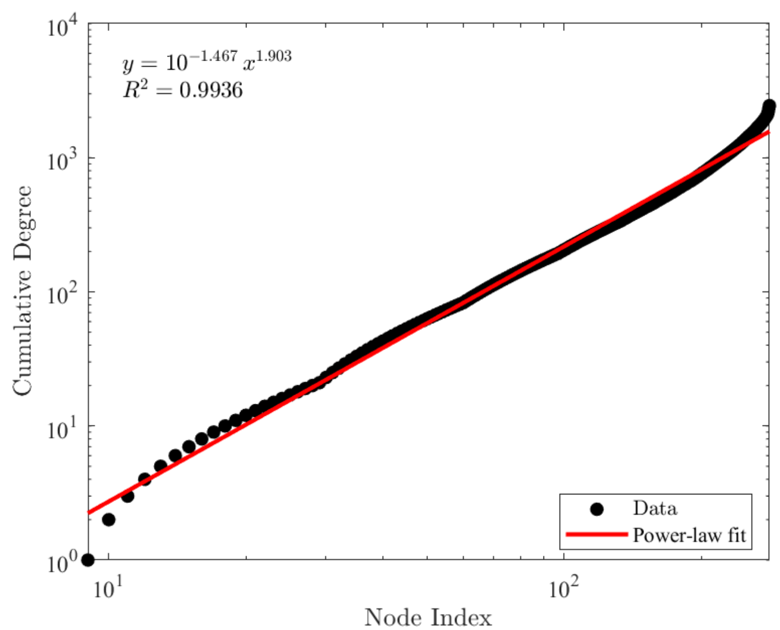

Figure 9 Node degree distribution follows a power-law trend

4.5.2.2 Weighted Degree Centrality

To quantify the structural prominence of contributing factors, I analyzed weighted degree centrality. In this study, I define importance as the extent to which a factor frequently co-occurs with other factors across accident reports. Unlike unweighted degree centrality, weighted degree centrality accounts for both the number of connections a node has and the strength of those connections. This measure identifies the factor that not only connects to many other nodes but also recurs in common accident patterns.

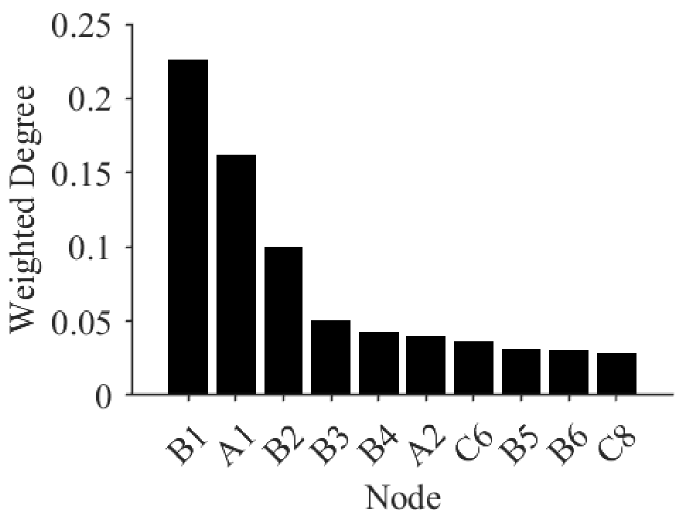

Figure 10 Weighted degree centrality of the ten highest ranked nodes

Figure 10 shows the top ten nodes ranked by weighted degree centrality. The nomenclature of each node follows the official NTSB finding code classification system and reflects the hierarchical structure defined in the NTSB database. Table 4 summarizes the top five nodes based on weighted degree centrality. For each node i, I calculated weighted degree centrality using Equation 4

$$S_i = \sum_{j \neq i} w_{ij} \qquad \text{Equation 4}$$

where s_i represents the weighted degree centrality of the node i, and w_ij represents the normalized edge between the nodes i and j. The normalized edge weight is calculated using

$$w_{ij} = \frac{c_{ij}}{\sum_{p<q} c_{pq}} \qquad \text{Equation 5}$$

where c_{ij} represents the number of Event IDs in which the nodes i and j appeared together. The denominator represents the total number of observed pairwise co-occurrence instances of all unique finding code pairs.

I describe high-centrality nodes as co-occurring with or bridging other findings within the same accident reports. Because the network is undirected and built from co-occurrence, this language is structural and should not be read as implying temporal sequence or causal direction.

Table 8 The top five nodes, their ranks, and significance in the analysis based on weighted degree centrality

Node ID	Nomenclature	Rank	Significance
B1	Personnel Issues-Action/decision-information processing/decision-Decision Making/Judgment-Pilot.	1	Reinforces its role as the network's primary hub. Its connections are both numerous and frequent.
A1	Aircraft- Aircraft Operations/Performance/Capability-Performance/ Control Parameters-Altitude- Not Attained/ Maintained	2	Indicates that altitude related findings are highly frequent and a central outcome of other failures
B2	Personnel Issues-Task Performance-Use of equipment/information-Aircraft Control-Pilot	3	Shows that errors in direct aircraft control are a consistent and influential factor.
B3	Personnel Issues- Psychological-Attention/Monitoring-Monitoring Environment-Pilot	4	The high weighted rank indicates that breakdowns in environment monitoring are a frequent part of the accident sequence.

B4	Personnel Issues-Task Performance-Planning/ Preparation- Flight Planning/navigation-pilot	5	The emergence of this factor in the top five of the weighted analysis is a key finding. It reveals that while planning errors may have fewer unique connections, they are highly frequent and contribute to accident pathways

The result reinforces the structural dominance of B1, A1, and B2. These factors not only connect to many other nodes, but they also frequently appear in the most common accident combinations. The emergence of B4 (Personnel Issues-Task Performance-Planning/preparation-Flight Planning/navigation) among the top five nodes is particularly noteworthy. Although it may have fewer unique connections than other nodes, it appears repeatedly in high-frequency accident patterns, indicating its consistent presence in CFIT operations in mountainous environments.

Additionally, the cumulative weighted degree distribution follows a power-law relationship $\log_{10} Y = 2.03 log_{10} X - 4.98$. This pattern further supports the presence of a scale-free structure in the weighted network. In practical terms, a small number of factors account for a disproportionately large share of co-occurrence strength. This finding suggests that safety interventions focused on these high-weighted nodes may provide greater leverage for reducing CFIT risk in mountainous operations.

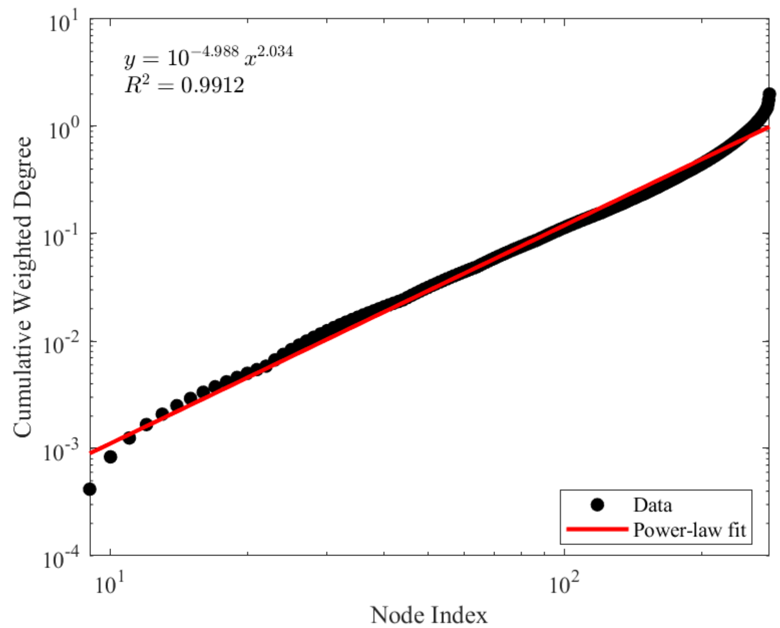

Figure 11 The weighted node degree distribution also follows a power law

4.5.2.3 Betweenness Centrality

My analysis identified the top five factors with the highest betweenness centrality in CFIT mountainous accidents. Betweenness centrality measures the extent to which a node lies on the shortest paths between other nodes in the network. In this study, a higher betweenness value indicates that a factor frequently connects otherwise separate groups of factors within the overall accident structure. Table 9 shows the top five nodes ranked by betweenness centrality, along with their nomenclature and interpretation.

Table 9 The top five nodes, their ranks, and significance in the analysis based on betweenness centrality metrics

Node ID	Nomenclature	Rank	Significance
B1	Personnel Issues-Action/decision-information processing/decision-Decision Making/Judgment-Pilot.	1	This node lies on the shortest paths among other contributing factors. Its high betweenness centrality indicates that pilot judgment frequently links distinct clusters of issues in CFIT mountainous accidents.

A1	Aircraft-Aircraft Operations/Performance/Capability-Performance/Control Parameters-Altitude-Not attained/maintained.	2	Altitude control frequently links judgment-related issues with operational outcomes, placing it in a structurally central position.
B2	Personnel Issues-Task Performance-Use of equipment/information-Aircraft Control-Pilot.	3	Aircraft control findings frequently co-occur with both decision-related and performance-related findings within the same accident report.
B3	Personnel Issue-Psychological-Attention/monitoring-Monitoring Environment-Pilot	4	Environmental monitoring frequently links cognitive factors with operational errors in the network structure.
A9	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Descent/approach/glide path	5	Descent and approach-related findings frequently co-occur with decision-related findings across accident reports.

The high betweenness centrality of these nodes indicates that they frequently connect multiple contributing factors within the network. Rather than appearing in isolation, these factors tend to co-occur across clusters of issues in accident reports.

For example, consider a small single-engine aircraft, such as a Cessna 172, flying over Colorado's Front Range at an altitude above 10,000 ft MSL. In these conditions, density altitude often exceeds 12,500 ft, climb rates may fall below 300 ft/min, and continuous turbulence requires sustained manual control. The resulting increase in workload, combined with mild hypoxia, can degrade pilot judgment (B1). In the network structure, B1 lies on many of the shortest paths connecting other contributing factors. When judgment deteriorates under these conditions, it frequently co-occurs with altitude-control errors (A1) and aircraft-control lapses (B2) in accident reports. This structural position explains why pilot judgment functions as a central connector among multiple contributing factors in CFIT accidents in mountainous terrain.

Figure 12 shows the distribution of betweenness centrality values across the network. The concentration of high betweenness in a small number of nodes reinforces the presence of structurally central factors that link multiple accident elements. These findings further support the identification of pilot decision-making, altitude management, and aircraft control as key structural connectors in CFIT accident patterns in mountainous terrain.

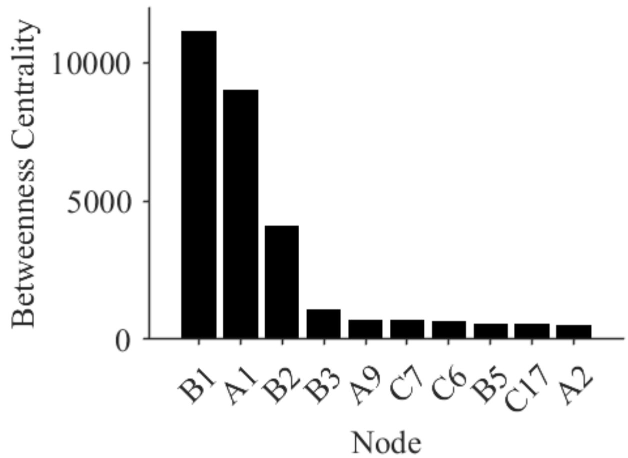

Figure 12 Frequency plot showing the betweenness centrality of nodes

4.5.3 Accident Reduction Strategies

The following nodes emerged as important across all centrality metrics (degree, weighted degree, and betweenness), thereby representing the highest-leverage intervention points: “Personnel Issues-Action/decision-information processing/decision-Decision Making/Judgment-Pilot” (B1), “Aircraft-Aircraft Operations/Performance/Capability-Performance/Control Parameters-Altitude-Not attained/maintained”(A1), “Personnel Issues-Task performance-Use of equipment/information-Aircraft Control-Pilot”(B2), and “Personnel Issue-Psychological-Attention/monitoring-Monitoring Environment-Pilot” (B3). Because these factors frequently co-occur with multiple other factors, they represent high-leverage intervention points within the structural accident network.

Based on these findings, I propose the following targeted strategies:

1. Mitigation of risks associated with (*Personnel Issues-Action/decision-information processing/decision-Decision Making/Judgement-Pilot- B1*): Enhanced cognitive training can improve pilot judgment under high workload conditions. Advanced simulator-based training should expose pilots to realistic mountainous terrain scenarios that require rapid decision-making under pressure. Integrating decision-support systems that provide structured, prioritized information may further assist pilots during high-demand phases of flight.
2. Addressing (*Aircraft-Aircraft Operations/Performance/Capability-Performance/Control Parameters-Altitude-Not attained/maintained-A1*): Structured training programs should emphasize altitude control in mountainous terrain and high-density altitude environments. A decision-support interface that continuously monitors altitude performance relative to terrain constraints may help pilots detect deviations earlier. Combining training with real-time performance feedback may reduce altitude-related errors.
3. The effective use of cockpit equipment (*Personnel Issues-Task performance-Use of equipment/information-Aircraft Control-Pilot-B2)*: Ergonomic interface design and improved human-machine interaction can reduce misinterpretation of cockpit information. Clear procedural guidance and standardized equipment-use protocols may enhance aircraft control performance, particularly during approach and descent in mountainous environments.
4. Enhancing pilots' attention and environmental monitoring capabilities (*Personnel Issue-Psychological-Attention/monitoring-Monitoring Environment-Pilot-B3*): Situational awareness training should emphasize terrain cues, weather

interpretation, and workload management in mountainous settings. Augmented visual and auditory alerts may support continuous environmental monitoring and reduce oversight during critical flight phases

4.5.4 Conclusion

Based on the network analysis of 287 accidents where mountain/hilly operations were causal/contributing factors, “Personnel Issues-Action/decision-information processing/decision-Decision Making/Judgment-Pilot” (B1), “Aircraft-Aircraft Operations/Performance/Capability-Performance/Control Parameters-Altitude-Not attained/maintained”(A1), “Personnel Issues-Task performance-Use of equipment/information-Aircraft Control-Pilot”(B2), and “Personnel Issue-Psychological-Attention/monitoring-Monitoring Environment-Pilot” (B3) consistently exhibited high structural centrality. Degree, weighted degree, and betweenness centrality indicate that these factors frequently co-occur with many other contributing elements and occupy central positions within the accident network.

The short average path length further suggests that contributing factors tend to appear in closely related patterns rather than in isolation. These findings do not imply direct causal propagation; rather, they highlight the structural prominence of specific cognitive and performance-related factors in CFIT accidents in mountainous areas.

This evidence supports targeted interventions focused on pilot decision-making, altitude management, aircraft control, and environmental monitoring. By prioritizing these structurally central factors, safety strategies can concentrate on areas most consistently associated with accident patterns. This approach provides a data-driven foundation for developing proactive decision-support systems and improving training methodologies. Future research may extend this framework by incorporating directed networks, probabilistic modeling, or integration with structured taxonomies, such as the Human Factors Analysis and Classification System (HFACS), to further refine the understanding of accident mechanisms in mountainous environments.

I applied complex network theory to NTSB accident data, using degree, weighted degree, and betweenness centrality to rank contributing factors. This approach depends on how narrative accident reports are coded into nodes and edges, and centrality ranking can shift with coding granularity and network-construction choices (directed or undirected). Network centrality captures the co-occurrence structure of contributing factors rather than causal ordering. Other analytical methods could serve the same risk-identification role and might prioritize a somewhat different set of information elements. For example, FMEA could rank contributing factors by severity, occurrence, and detectability whereas Bayesian network models could model conditional dependencies among factors. Furthermore, the critical nodes identified through the network analysis were not independently validated. Engaging subject matter experts, such as experienced mountain flying instructors, examiners, or accident investigators, to review the critical nodes identified by the centrality analysis would provide an operational check on the data-driven ranking and strengthen confidence in the selected information priorities. The proposed framework does not depend on complex network theory; it requires only a defensible, data-driven method for identifying context-critical information. Complex network analysis was therefore one instantiation of the risk-identification stage, and future applications of the framework should compare complex network analysis with alternative methods and submit the resulting information priorities for expert review.

4.6 Mapping Critical Factors to the Generic Interface

The identified critical factors (nodes) inform the prioritization of information elements within the generic interface described in Chapter 3. The complex network analysis revealed four dominant nodes based on degree, weighted degree, and betweenness centrality. “Personnel Issues-Action/decision-information processing/decision-Decision Making/Judgment-Pilot” (B1), “Aircraft-Aircraft Operations/Performance/Capability-Performance/Control Parameters-Altitude-Not attained/maintained (A1),” “Personnel Issues-Task performance-Use of equipment/information-Aircraft Control-Pilot”(B2), and “Personnel Issue-Psychological-Attention/monitoring-Monitoring Environment-Pilot” (B3). These nodes consistently occupied structurally central positions within CFIT mountainous accident patterns.

4.6.1 Mapping Logic

The spatial organization of the interface follows the generic design framework established in Chapter 3. That framework defines fixed display zones based on human factors principles, cognitive workload management, and FAA display guidance. The network analysis does not determine the placement of these zones. Instead, it identifies which mission-specific factors should populate the high-salience regions within the predefined structure. This distinction preserves layout consistency while enabling mission-tailored prioritization.

To translate the network findings into design requirements, I applied a functional alignment process. I categorized each high-centrality node according to its operational function (decision-making, altitude management, aircraft control, or environmental monitoring). I then mapped those functional categories onto the corresponding zones defined in the generic framework. Factors directly related to aircraft performance and terrain clearance (A1) were assigned to visually dominant areas. Cognitive and judgment-related factors (B1) were supported through predictive metrics and alert panels. Monitoring-related factors (B3) were reinforced through terrain visualization and environmental displays. This structure mapping ensures that the interface remains architecturally generic while being contextually tailored. Table 10 shows how centrality nodes were mapped on the interface.

Table 10 Each key node is mapped to interface content

Central Node	Network Interpretation	Interface Implication	Interface content
B1: Decision Making/Judgment-Pilot	Central cognitive factors across accident patterns	Support faster decision-making under a workload	Advisory panel, decision cue, supplementary panel
A1: Altitude not attained/maintained	Recurrent aircraft-performance/altitude control issue	Make terrain-relative altitude highly visible	Terrain clearance, altitude margin, vertical profile.
B2: Aircraft Control-Pilot	Links cockpit information to control performance	Support aircraft state monitoring	ROC, air speed, flight path deviation.
B3: Monitoring Environment-Pilot	Reflects a breakdown in environmental monitoring	Improve environmental awareness	Terrain and hazard cues.

The tailored interface shown in Figure 13 directly instantiates the generic framework presented in Chapter 3. The “main content” region in the generic structure corresponds to the central terrain-and-flight-path visualization area in the tailored interface. The “critical alert panel” aligns with Area 7 (Critical Alerts/Notification), while the “Status Board Panel” maps to Area 6 (Global Status Summary). The “supplementary panel” corresponds to Area 5 (Upcoming Segment Metrics). The navigation sidebar and footer retain their structural roles as defined in the generic framework. This tailored design preserves the hierarchical architecture of the generic framework while replacing mission-specific content with content relevant to mountainous operations.

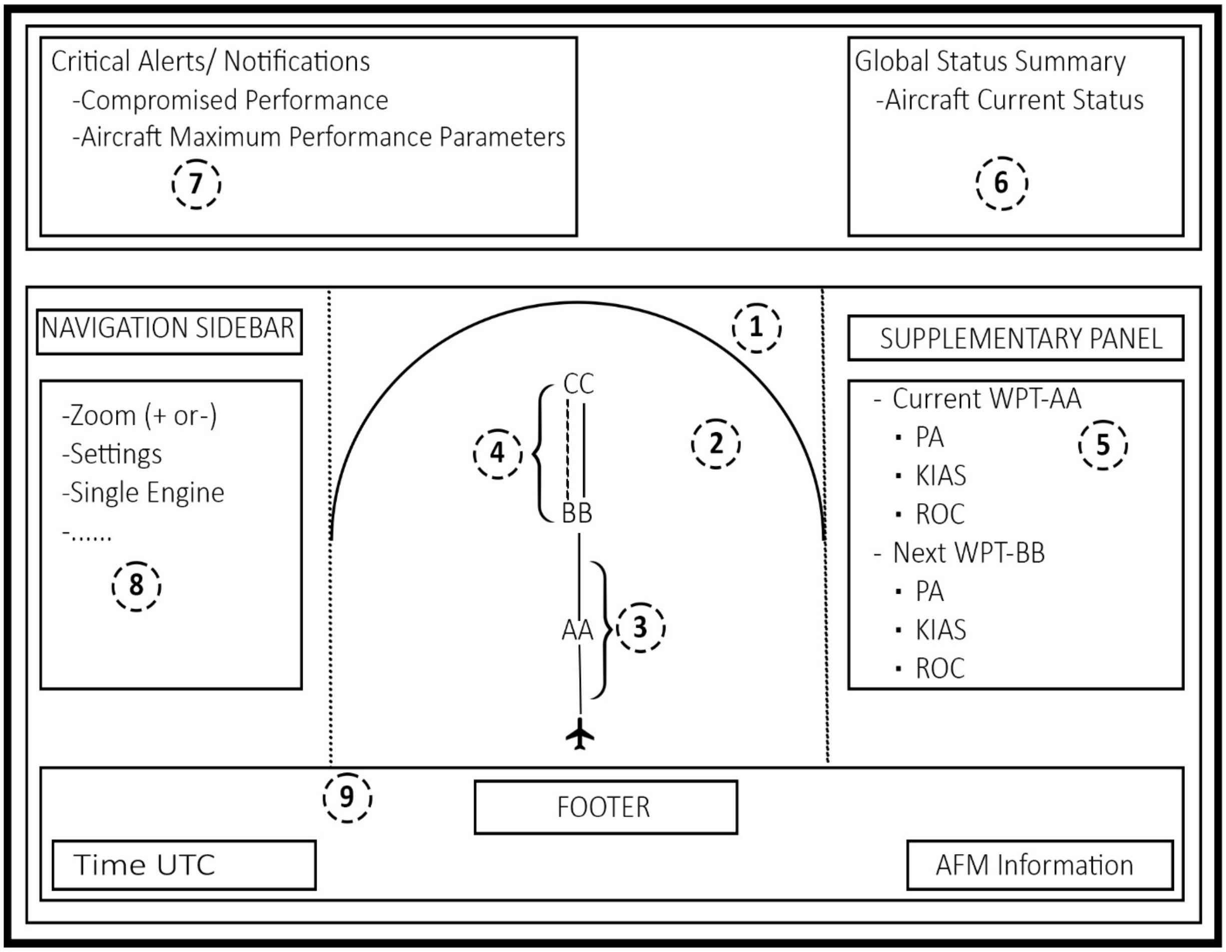

Figure 13 Low-fidelity tailored interface wireframe derived from the generic interface architecture

4.6.2 Adaptive Behavior of the Interface

The interface operates as a rule-based adaptive system. The spatial layout remains fixed in accordance with the generic framework described in Chapter 3. Adaptation occurs at the level of information content and visual emphasis rather than structural rearrangement.

As the aircraft altitude changes, the interface dynamically updates the corresponding Aircraft Flight Manual (AFM) performance values, such as rate of climb (ROC) and airspeed, for the current altitude. It also retrieves and displays AFM performance parameters associated with the planned altitude at the next waypoint extracted from the active flight plan. Thus, the interface continuously aligns certified performance data with the aircraft operational context.

In addition to updating AFM values, the system applies a user-defined degradation threshold relative to baseline performance (e.g., 50% reduction from sea-level capability). When performance falls below this threshold, the interface increases visual salience by providing a “compromised performance” cue. This escalation does not alter the layout; it modifies color and emphasis to draw attention to reduced performance margins.

Terrain visualization remains continuously visible, similar to conventional glass cockpit systems. However, by presenting AFM-derived performance information alongside terrain context and waypoint altitude requirements, the interface supports anticipatory situation awareness without introducing layout variability.

Thus, the interface adapts through contextual data updating and threshold-based visual escalation while preserving spatial consistency and cognitive stability.

4.6.3 Tailored design

The tailored interface is organized into nine distinct areas; each serves a different function and is mapped to the identified critical nodes to ensure comprehensive risk mitigation.

1. Area 1 (Track Up Compass): Provides fundamental heading information, supporting B2 (Aircraft Control) and general situational awareness.
2. Area 2 (DEM-based terrain Map): Presents a digital elevation model (DEM)-based terrain map, with terrain above the altitude of the aircraft highlighted in red. It also

shows the flight plan path between Waypoints (e.g., AA, BB, CC). This area can be toggled between full and outlined views via the control in Area 8. This is crucial for supporting B3 (Environmental Monitoring) by enhancing awareness of terrain and for aiding B1 (Decision Making) by providing topographical context.

3. Area 3 (Normal Flight Plan Path Segment): Shows the flight plan path (e.g., between waypoints AA and BB) using a solid black line for normal performance, supporting B2 (Aircraft Control)
4. Area 4 (Performance-Compromised Flight Plan Path Segment): Shows a solid red line (e.g., between BB and CC) to indicate a pre-flight performance compromise on that segment. The path will blink red once the aircraft initiates that segment. This directly addresses A1 (Altitude-Not Attained/Maintained) by making performance limitation salient and supports B1 (Decision-Making)
5. Area 5 (Upcoming Segment Metrics): Lists upcoming segment metrics to support proactive planning. This aids B1 (Decision making) and prepares for B2 (Aircraft Control).
6. Area 6 (Real-time flight parameters/Global Status Summary): Shows real-time flight parameters. This is vital for B2 (Aircraft Control), B1(Decision Making), and B3 (Environment Monitoring). Interface-B would provide comprehensive parameters, while Interface-A would provide essential parameters.
7. Area 7 (Critical Alerts/Notifications & AFM Maximum Performance Parameters): Displays a prominent red header warning when a published altitude or performance threshold is crossed (mirrors violations in Area 4). In addition, it also houses the AFM’s maximum performance parameters. This is critical for

addressing A1 (Altitude-Not Attained/Maintained) and Supporting B1 (Decision Making). Interface-B ensures comprehensive multi-modal alerts.

8. Area 8 (Navigation Sidebar-Zoom, Settings, Mission-Specific buttons): Contains zoom, display settings, and mission-specific buttons (for example, single-engine performance settings on multi-engine aircraft) that let the pilot customize terrain detail and performance overlays. This supports B1 (Decision Making) and B3 (Environmental Monitoring) by allowing adaptable information display.
9. Area 9 (Footer-Time UTC and AFM Information): Shows the time and active AFM version, ensuring the correct manual reference without diverting attention from primary flight data, subtly supporting B1 (Decision Making) and B2 (Aircraft Control)

A high-fidelity wireframe of the concept mentioned above is shown in Figure 14. It illustrates the CFIT mountainous operation interface populated using the proposed fixed layout and context-sensitive framework. The interface is organized into four primary display regions and one footer/status region. The header bar contains critical alerts, aircraft performance limits, and current aircraft status. In the left portion of the header, the interface displays a compromised-performance notification, indicating that the aircraft's performance has degraded below a user-defined threshold (here, I used 50%). In this example, the warning indicates compromised performance from EVERR to JASAT. This warning will be displayed if the aircraft has surpassed the user-defined threshold. Below the alert, the aircraft performance-limit panel provides the maximum attainable performance for that aircraft type as specified in the AFM. These values are displayed in yellow/amber to indicate caution-level information and to distinguish them from routine status information.

The top-right side of the header bar displays the current aircraft state. In this example, the aircraft is at an altitude of 9,000 ft, traveling at 80 knots, and climbing at 50 ft/s, as indicated by the upward arrow next to the vertical speed value.

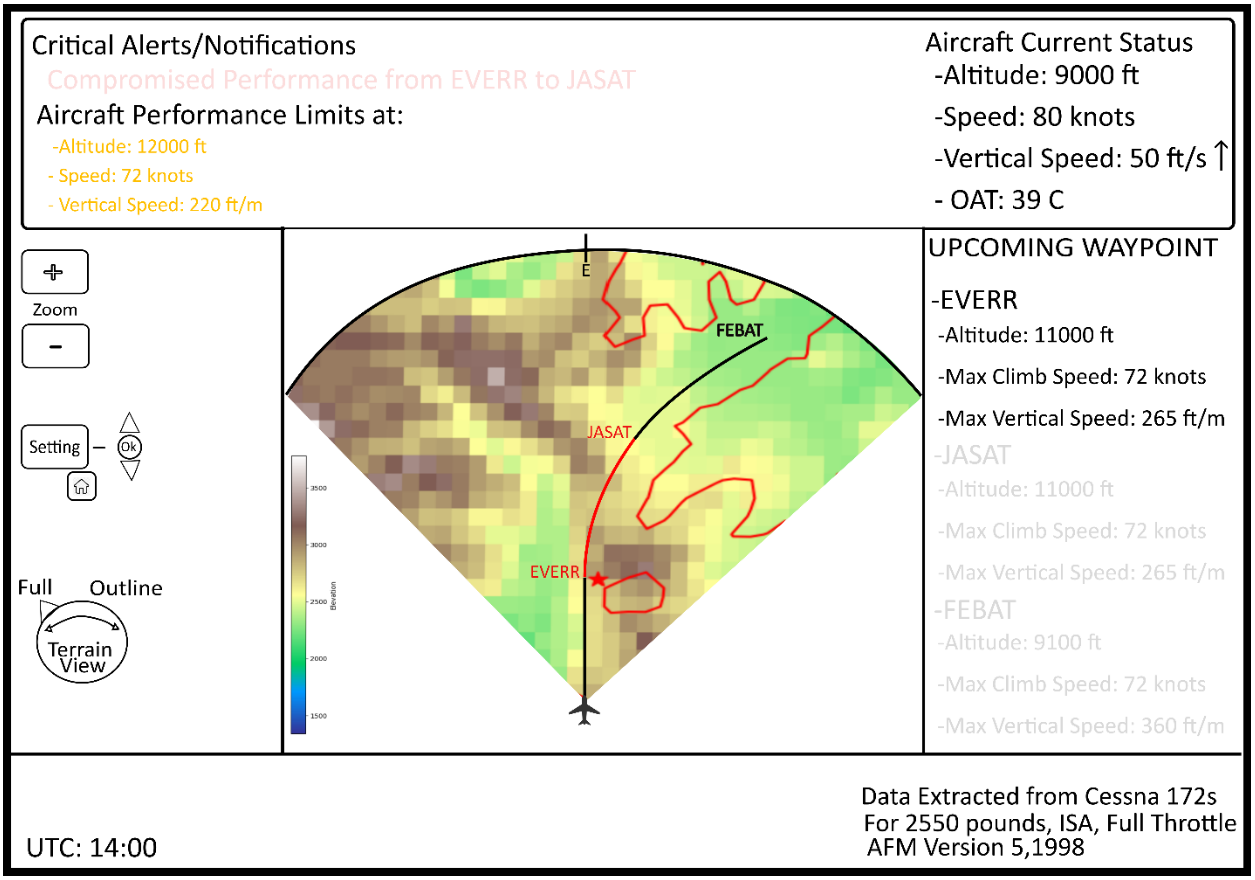

Figure 14 Initial high-fidelity wireframe of the interface showing how information elements are displayed

The outside air temperature is also displayed as 39°C. This placement allows the pilot to compare the current aircraft condition with the aircraft performance limit without searching across multiple displays.

The navigation sidebar on the left side of the interface contains user-input controls, including zoom settings and view-selection buttons. These controls are separated from the primary operational display to preserve the stability of the main content region and to prevent configuration functions from competing with safety-critical information. The main region presents the terrain profile and route geometry. In this example, the terrain map displays the aircraft position, the route toward upcoming waypoints, and terrain elevation using the embedded color legend. This region is reserved for mission-critical information because terrain awareness and climb performance assessments are central to CFIT risk management in mountainous operations. By presenting terrain, route, aircraft status, and position in the same region, the interface reduces the need for the pilot to mentally integrate these elements from separate displays.

The supplementary panel on the right provides upcoming waypoint information. For the next waypoint, EVERR, the displays show the required altitude constraint of 11,000 ft, along with the maximum climb speed and vertical speed information available to the aircraft. After the aircraft passes EVERR, the next waypoint information is updated so the pilot receives context-specific guidance for the next segment of the flight. Future waypoint information is visually de-emphasized to avoid competing with the active waypoint. The footer/status bar provides supporting information that does not require continuous primary attention. The left side displays the UTC time, while the right side identifies the source and assumptions used for the aircraft performance data. In this example, the performance values were extracted from the Cessna 172 data for a 2,550-lb aircraft, under ISA conditions, at full throttle, and using AFM Version 5 (1998).

Including these assumptions increases transparency and helps the pilot understand the basis for the displayed performance limits.

The compromised performance alert is generated directly from the AFM performance tables rather than from a computed performance model. The AFM publishes rate of climb and speed values at a series of altitudes. The interface compares the AFM rate of climb and speed value, corresponding to the aircraft current altitude against the published sea-level values, and the alert is displayed when the tabulated performance at the current altitude falls below a user-defined fraction of the sea level value (50% in this study). The alert is therefore driven solely by the aircraft altitude relative to AFM performance tables. It appears whenever the aircraft crosses the threshold, and remains displayed while the condition persists, independent of flight path accuracy or pilot inputs.

Because the alert depends only on altitude relative to AFM performance data, it would also be displayed during a nominally flown approach. This behavior is intentional as it communicates that the available climb performance margin is reduced before any deviation or hazard develops, supporting proactive decision making rather than reacting to an already unsafe state.

The customized interface works in tandem with real-time terrain data and aircraft performance system (AFM-derived) metrics to deliver a multi-layered view of the aircraft status.

5 Prototype Design

Modern flight simulation platforms such as X-Plane 12 by Laminar Research offer high-fidelity aerodynamic modeling, global scenery rendering, and realistic avionics simulation. However, the closed nature of built-in avionics displays, such as the Garmin G1000 integrated flight deck emulation, limits researchers' and flight organizations' ability to extract, augment, and present telemetry data in formats tailored to specific experimental and pedagogical requirements. A pilot in training, for instance, may benefit from an external display that simultaneously shows real-time speed, altitude, vertical speed, terrain proximity (color-coded to EGPWS standards), the active flight plan with cross-track deviation, and AFM-derived climb performance envelopes. These capabilities are not natively combined in any single avionics page within the simulator.

This chapter presents the design, architecture, and implementation of a four-component software system developed to test whether an integrated telemetry panel could reduce pilot cognitive workload by consolidating terrain awareness, flight plan tracking, and climb performance information into a single unified display. The system consisted of two custom X-Plane plugins that ran inside X-Plane 12, a flight plan bridge plugin, and an external PySide6 graphical interface. The design philosophy is grounded in human factors principles, where the display consolidates information that the pilot would otherwise need to cross-reference across multiple avionics pages, thereby reducing visual scanning demands and working memory load.

The remainder of this chapter is organized as follows: Section 5.1 provides background and context. Section 5.2 presents system architecture and data flow. Sections 5.3 through 5.5 describe each component in detail. Section 5.6 discusses the terrain awareness colorization algorithm. Section 5.7 covers waypoint management and active leg tracking. Section 5.8 describes the graphical user interface design.

5.1 Background and Context

X-Plane 12 is a blade-element theory flight simulator widely used in academic flight dynamics research, pilot training, and human factors studies. The simulator exposes thousands of internal variables (datarefs) via a C-written plugin Software Developer Kit (SDK) and supports third-party plugins through the XPPython3 framework, which embeds a Python interpreter within the simulator's process. This plugin architecture permits researchers to read and write simulator state at the flight-loop call-back rate, typically 20-60 Hz, without modifying the simulator source code.

The X-Plane 12.1 update introduced a new Flight Plan Application Program Interface (API) that provides structured access to pilot-entered FMS routes, including primary, approach, and temporary plan kinds. This API supersedes the legacy single-plan interface and enables more faithful extraction of procedures as programmed in the G1000 avionics suite. The present work leverages both the legacy and 12.1 APIs, automatically falling back to the newer interface when the older interface is unavailable.

5.1.1 Related Work

Several open-source projects have addressed parts of the telemetry extraction problem. The XPlaneConnect (XPC) library developed by NASA provides a User Datagram Protocol (UDP)-based interface for reading and writing X-Plane datarefs from external applications. The prototype developed in this study uses an XPC as the primary telemetry transport for airspeed, altitude, vertical speed, heading, and position. Other projects, such as X-Plane Map and various glass-cockpit overlays, provide map-based flight tracking but typically do not integrate real-time terrain probing, EGPWS colorization, or AFM performance data into a single display. Crucially, none of these tools have been designed with an explicit objective of cognitive load reduction or evaluated within the theoretical frameworks of Endsley's SA or Wickens' Multiple Resource Theory. Tang et al. (2024) compared the cognitive effects of steam gauge versus G1000 glass-cockpit displays among novice pilots and found that display integration significantly affected both subjective workload (NASA-TLX) and physiological indicators (EEG, HRV), highlighting the importance of display design in workload management. The FAA’s guidance on flight deck design similarly emphasizes that information consolidation, appropriate use of color coding, and minimization of required cross-referencing are key strategies for reducing pilot workload. The present system is motivated by these findings and seeks to test whether a purpose-built external display that consolidates terrain, navigation, and performance data can achieve measurable reductions in workload under demanding operational conditions.

5.2 System Architecture and Data Flow

The system comprises four software components that cooperate through three distinct data exchange mechanisms. Figure 15 illustrates high-level architecture and the flow of data between components.

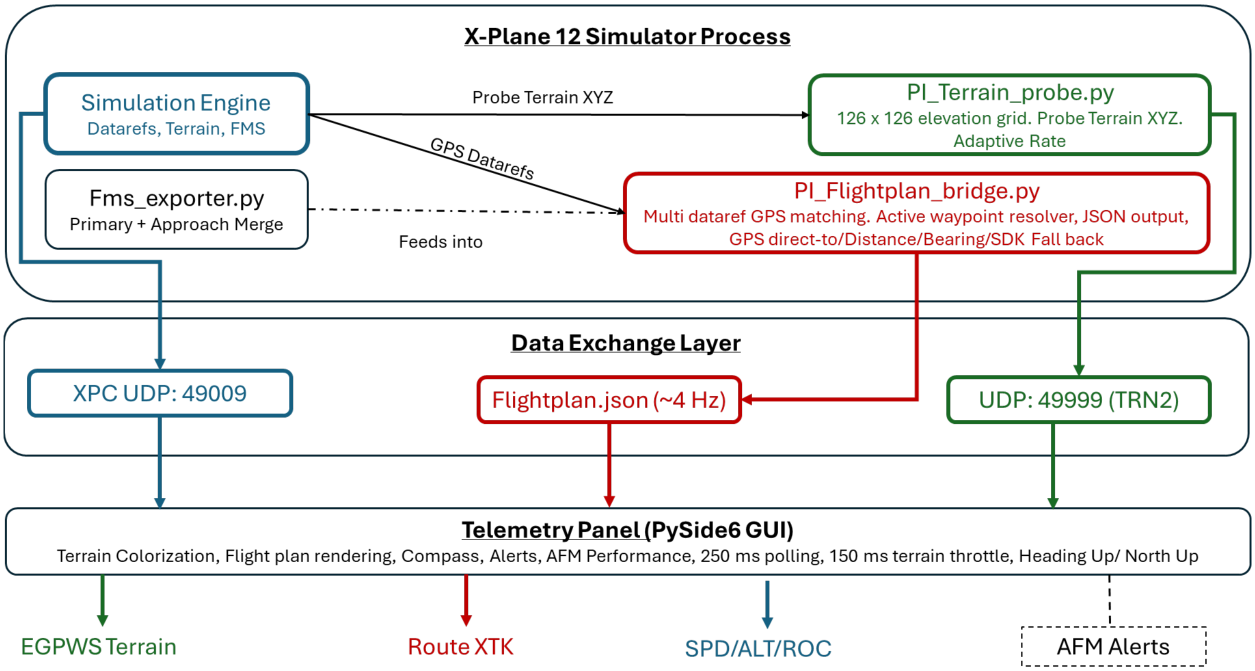

Figure 15 System Architecture and data flow. Three plugins execute within the X-Plane 12 process; the telemetry panel runs externally and consumes data through UDP, JSON, and XPC channels

The color coding in the diagram reflects the data lineage: green traces the terrain data path from the probe plugin through the TRN2 UDP streams to the telemetry panel. The crimson traces the flight plan data from the bridge plugin through the JSON file to the route and XTK display. The blue trace shows the live data reference path from the simulation engine through XPlaneConnect to the numerical readout. The black traces the FMS and AFM performance data to the alerts panel.

5.2.1 Data Exchange

The three data exchange mechanisms are designed to be loosely coupled and independently verifiable.

1. UDP Terrain Stream (TRN2 protocol): The terrain probe plugin transmits a binary UDP datagram to localhost: 49999 at an adaptive rate. Each datagram carries a structured header containing the frame dimensions (N×N grid), the geographic

center (latitude, longitude), a pixels-per-nautical-mile scaling factor, and the minimum and range of terrain elevation values in meters. The payload is an N×N array of unsigned 8-bit values representing normalized terrain height.

2. JSON Flight Plan File: The flight plan bridge plugin writes a JSON file at approximately 4 Hz containing the full set of FMS waypoints, the determined active waypoint index and identifier, GPS diagnostic data ref values, and the method used to resolve active leg. The telemetry panel polls this file at 250 ms intervals and detects structural and active-leg changes by comparing signatures.
3. XPlaneConnect (XPC) Datarefs. The telemetry panel queries live datarefs over UDP using the NASA XPlaneConnect Library, extracting indicated airspeed, barometric altitude, vertical speed, magnetic heading, geographic position, altitude above ground level, VNAV targets, and altitude preselect values.

5.2.2 Timing and Synchronization

The system does not require frame-level synchronization between components. The terrain probe runs within X-Plane's flight loop at an adaptive rate (0.25-1.5 seconds per frame, governed by ground speed and turn rate). The flight plan bridge runs at a fixed 0.25-second interval. The telemetry panel's main polling timing fires every 250 ms, and the terrain stream handler throttles incoming frames to a minimum of 150 ms separation. This design ensures that all components remain responsive without excessive CPU utilization.

5.3 Terrain Probe Plugin

This plugin generates a real-time elevation raster of the terrain surrounding the aircraft and transmits it to the telemetry panel via UDP. The plugin constructs a default

126 × 126 elevation grid centered on the current position of the aircraft. The grid covers a configurable range of ± 40 nautical miles in both the north-south and east-west directions. For each grid cell, the plugin invokes X-Plane's *probeTerrainXYZ* function, which performs a downward ray-cast from a reference height (10,000 meters above the local origin) and returns the terrain elevation in meters at that horizontal position.

The raw elevation values are normalized to the range [0, 255] using the minimum and maximum of the sampled elevations. The header is packed using Python's *struct* module in little-endian format, with the magic bytes TRN2, followed by the version, grid dimensions, geographic center, pixels-per-nautical-mile (PPNM), minimum elevation, and elevation range.

One of the features in this plugin is adaptive transmission rate, which balances terrain freshness against computational cost. The plugin tracks the aircraft ground speed and heading turning rate, and computes the next transmission period as the more restrictive of the two constraints:

a) The time for aircraft to traverse a configurable fraction of a grid pixel (default 0.4 pixels), and
b) The time for the heading to change by a configurable threshold (default 1.0 degree).

The resulting period is clamped to [0.25, 1.50] seconds. When the aircraft is parked with wings level, the rate defaults to the maximum period to conserve CPU resources.

5.4 FMS Exporter Plugin

The tool reads the current G1000 flight plan from X-Plane, using the new X-Plane 12.1 flight plan system when available, and, otherwise, the older FMS-reading functions. When both the primary and approach flight plans contain entries, the plugin merges them, preferring approach altitude constraints where the ident and position match. Approach-only fixes not present in the primary plan are appended to the merged list. This strategy ensures that altitude constraints from published procedures are preserved even when the primary plan does not encode them. To avoid excessive disk I/O, the plugin computes a signature tuple of (ident, latitude, longitude, altitude) for each entry and compares it against the previous export. If the signature is unchanged and fewer than 0.5 seconds have elapsed, the write is suppressed.

5.5 Flight Plan Bridge Plugin

The flight plan bridge is the most complex of the three plugins. Its primary responsibility is to determine which waypoint in the active flight plan is the current “TO” waypoint, i.e., the fix ahead of the aircraft toward which it is currently navigating, and to export this information along with the full waypoint list.

Determining the active waypoint is non-trivial because X-Plane’s SDK does not provide a single authoritative dataref that reliably identifies the TO waypoint across all avionics modes and flight plan configurations. The bridge implements a cascade resolution strategy with four priority levels, illustrated in Figure 16

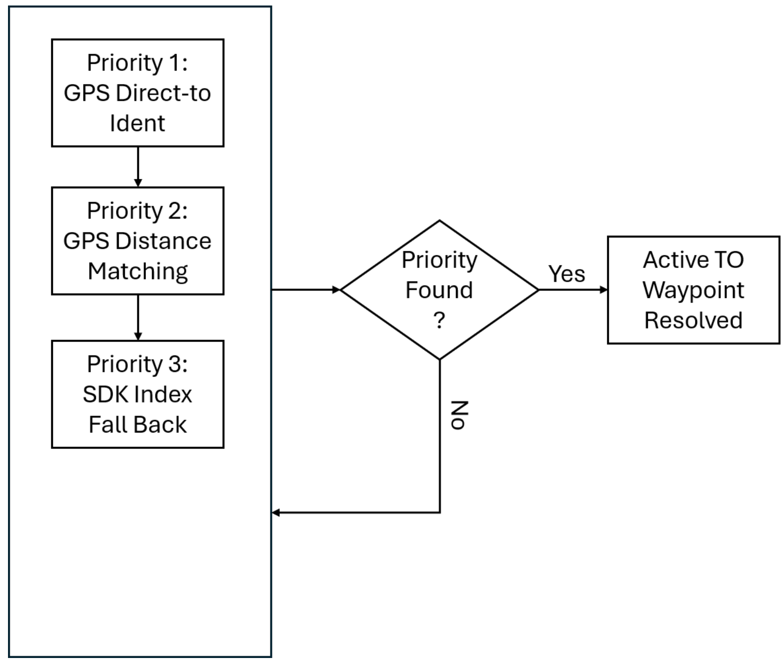

Figure 16 Priority cascade used by the flightplan bridge to resolve the active waypoint

At the highest priority, the bridge queries the Global Positioning System (GPS) Direct-To navaid reference. If this resolves to a valid identifier, the waypoint list is searched for a matching fix. If Direct-to is unavailable, the bridge reads four GPS distance datarefs and computes the great-circle distance from the aircraft to each waypoint, selecting the fix whose distance best matches the GPS-reported values within a 0.5 NM tolerance. If distance matching fails, three GPS bearing datarefs are probed and compared against computed bearings with a 15-degree tolerance. As a final fallback, the SDK reads the displayed FMS entry index and resolves it to the nearest valid fix. This multi-strategy approach was necessitated by empirical observations during development: no single GPS dataref is consistently populated across all flight phases, and by avionics configurations in X-Plane 12.

Each JSON export includes a diagnostic block that contains the raw values of all probed GPS datarefs, the selected resolution method, and the SDK's active index. This diagnostic output proved invaluable during iterative development and enabled post hoc analysis of waypoint resolution accuracy.

5.6 G1000 Terrain Colorization

The telemetry panel implemented a Garmin G1000 Terrain Proximity-style colorization method to display terrain clearance relative to the aircraft's current altitude. I did this so participants could rely on their natural interpretation, which required building an interface they are already familiar with. The method converted the greyscale terrain raster into a three-color terrain display using the same clearance logic used in G1000-style terrain proximity presentations: red for terrain above or within 100 ft below the aircraft, yellow for terrain between 100 ft and 1000 ft below the aircraft, and black for terrain more than 1000 ft below the aircraft.

The colorization process used four steps. First, the algorithm converted the 8-bit greyscale terrain pixels back to absolute terrain elevation in meters using the minY and rngY parameters transmitted in the TRN2 terrain header. Second, when above-ground-level data were available, the algorithm applied a local bias correction to reduce mismatch between the terrain raster and the simulator's AGL dataref at the center pixel. Third, the algorithm computed terrain clearance in feet as the difference between aircraft mean sea level altitude and terrain elevation. Fourth, the algorithm discretized the clearance values into three terrain-clearance bands and mapped each band to a precomputed RGB lookup table.

Table 11 Color coding schema and clearance level definition in the interface

Clearance Range (ft)	**Color**	**RGB Value**	**Interpretation**

≤ 100	High-density Red	(220,0,0)	Terrain is within or less than 100 ft below your aircraft
100 to 1000	Yellow	(220,180, 0)	Terrain is greater than 100 ft but less than 1000 ft below your aircraft
> 1000	Black	(0,0,0)	Terrain is greater than 1000 ft below your aircraft

The interface updated the terrain colorization at the terrain-frame rate, up to approximately 7 Hz, on a 126 × 126 terrain grid. The implementation used NumPy vectorized operations rather than per-pixel Python loops. The np.digitize function assigned all 15,876 terrain-clearance values to clearance bands in one operation, and lookup-table indexing converted the band indices into the final RGB terrain image. This approach allowed the terrain display to update within the prototype's real-time display requirements.

5.7 Waypoint Management

The *WaypointManager* class maintains an ordered list of Waypoint dataclass instances and tracks the index of the current TO waypoint *(active_to_idx).* The design follows the G1000 convention: the active leg runs from *waypoints[to_idx - 1] to waypoints [to_idx]*. When *to_idx == 0* (i.e., the aircraft is navigating directly to the first waypoint), the FROM endpoint is the aircraft's current position rather than a stored waypoint.

The flight plan watcher parses the JSON file produced by the bridge plugin and extracts segments to present only the relevant portion of the FMS procedure. A full instrument approach procedure loaded in the G1000 may contain eight or nine waypoints spanning the initial approach fix (IAF), intermediate fixes, the final approach fix, the runway threshold, missed approach waypoints, and holding fixes. The parser identifies key structural markers (e.g., EVERR as the IAF, RW27 as the runway, TEPRE as the missed approach climb fix) and selects either the approach segment or the missed approach segment based on which waypoint the bridge reports as active.

Cross-track deviation (XTK) is computed in the terrain widget's paint event using a projection of the aircraft's screen position onto the perpendicular of the active leg line. The perpendicular distance is converted to nautical miles using the screen pixels-per-nautical-mile scaling factor. The result is displayed as a magenta-bordered text indicator above the ownship symbol, formatted as "XTK 0.3 L" or "XTK 1.2 R", and is suppressed when the deviation is below 0.05 NM or above 10 NM.

5.8 GUI Design

The telemetry panel GUI is built with PySide6 (Qt 6 for Python) and is organized as a layered full-screen display with semi-transparent overlays. The visual design is informed by cognitive workload reduction principles: information that a pilot would normally need to gather from three or more separate avionics pages (terrain map, flight plan, performance data) is consolidated into a single display, reducing the visual scanning and working memory demands identified by Wickens as primary sources of task interference in multi-display cockpits.

The main window contains three visual layers. The bottom layer is the terrain widget, which renders the G1000 -colored terrain raster, the compass rose, the flight plan route, and the ownship symbol. The middle layer is a transparent overlay that carries the telemetry readouts (speed, altitude, vertical speed), the alerts panel, the waypoint sidebar, the range controls, and the G1000 legend. The top layer handles user interaction (keyboard shortcuts and button clicks).

The speed, altitude, and rate-of-climb readouts are displayed in a semi-transparent black panel in the upper-right corner. Values are quantized to reduce visual jitter: altitude to the nearest 10 feet, vertical speed to the nearest 100 ft/min with a 50 ft/min deadband, and airspeed displayed unquantized. A monospaced font (Consolas) is used for numerical readouts to prevent horizontal shifting as digits change.

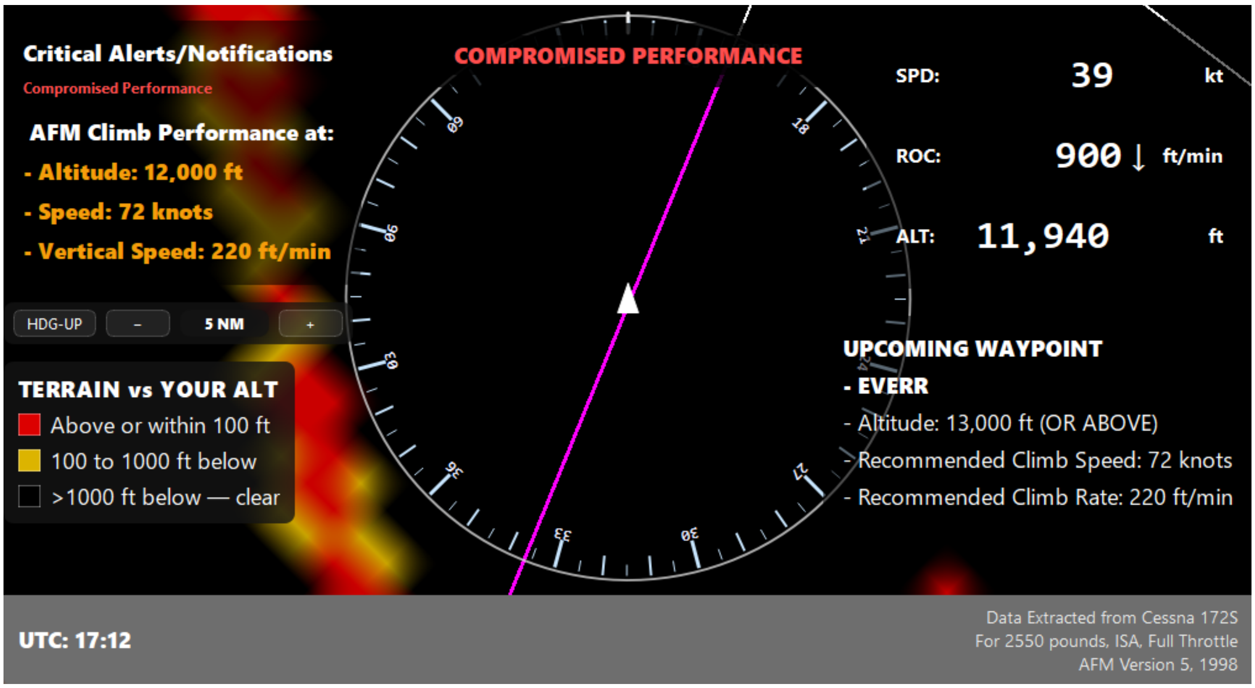

Figure 17 The final prototype of the interface

The alerts panel displays recommended climb performance values interpolated from the Cessna 172S Approved Flight Manual (AFM) climb table. The table encodes best-rate-of-climb speed and rate of climb at 1,000-foot altitude increments from sea level to 12,000 feet under ISA conditions at maximum gross weight (2,550 pounds) with full throttle. A “Compromised Performance” alert is triggered when the expected rate of climb at the current altitude falls below 50% of the sea-level rate.

The compass is rendered as a semi-arc with 5-degree tick marks, 10-degree intermediate ticks, and 30-degree labeled cardinal and ordinal headings. In heading-up mode (default), the compass rotates to keep the aircraft's magnetic heading at the top of the display. Heading values are smoothed using an exponential moving average (alpha = 0.2) to reduce visual jitter.

The terrain display range is adjustable through five discrete steps (2, 5, 10, 15, and 20 NM). The range is automatically clamped to the maximum that the terrain raster can support given its pixel dimensions and the pixels-per-nautical-mile scaling factor. Keyboard shortcuts provide rapid range adjustment during flight.

6 Method and Results

This chapter presents the methodology and results of the evaluation of the tailored cockpit interface developed in this dissertation. It describes experimental methodology, including the study design, ethical approval, simulator setup, participant instrumentation, flight scenario, data sources, sample size consideration, and statistical analysis strategy. This chapter also presents the results from subjective, physiological, flight performance, and visual attention analysis. The study evaluates whether the tailored interface supported pilot information processing during a simulated instrument approach in mountainous terrain. The experiment compared a baseline group, which completed the approach without the tailored interface, with a tailored-interface group, which completed the same approach with the interface available. This comparison allowed the study to examine whether the interface influenced perceived workload, situation awareness, physiological response, visual attention, and flight-path performance. The methodology section explains the data collection process and provides reasoning for the statistical methods used in the analysis. The result section then reports the findings for each outcome category. Together, these sections provide the empirical basis for assessing whether the tailored interface improved pilot support during a CFIT-relevant approach scenario.

6.1 Experimental Design

This study used a between-subjects, sequential cohort experimental design to compare pilot response and task performance across two interface conditions: baseline and tailored. Each participant completed one interface condition, which avoided learning and carryover effects. Both groups completed the same simulated RNAV RW 27 approach using the same VR simulator, control column, thrust pedals, and instrumentation. The only experimental difference was the interface condition: the baseline group completed the approach without the tailored interface, whereas the tailored interface group completed the approach with the tailored interface available. The inceptors (control column and thrust pedal) were force-feedback, i.e., the pilot could feel the real feedback. The force-feedback controls were used to preserve realistic control feel during the approach. The interface condition served as the independent variable. The dependent measures were subjective workload, perceived situation awareness, heart rate variability, pupil response, visual attention, and flight path performance. The analysis treated the two groups as independent samples because different participants completed the baseline and tailored-interface conditions.

This design aligned with the purpose of this study, which aims to evaluate whether a designed information presentation could support pilot information processing in a high-demand operational context. The experiment examined whether the interface helps pilots recognize, integrate, and use critical information during an IFR approach in mountainous terrain. Auburn University approved the baseline data collection under IRB protocol STUDY00000501 and the tailored-interface data collection under IRB protocol STUDY00001103. Before each session, the researcher explained the study purpose, the simulator task, the equipment, the data collection procedure, and the participants' rights. Participants completed the required consent process before the experiment began. The study allowed participants to stop the session at any time if they felt uncomfortable or no longer wanted to continue.

6.1.1 *Participant Demographics and Sample Size*

I recruited participants through flyers and email invitations. Flyers were distributed across different departments at Auburn University, and email invitations primarily targeted the School of Aviation. Eligible participants were pilots with at least basic instrument flight knowledge, defined as holding an instrument rating or having completed instrument training. Data collection occurred in two phases under separate IRB protocols: the baseline group completed the experiment in October 2025 (STUDY00000501), and the tailored interface group completed the experiment in May 2026 (STUDY00001103)

The baseline group initially consisted of 32 participants, of whom 22 were included in the analysis (16 male, 5 female, 1 preferred not to answer; 20 aged 18-29 and 2 aged 60+). The tailored interface group consisted of 12 participants (10 male, 2 female, 11 aged 18-29 and 1 aged 30-39) who had instrument approach experience. I used the minimum total flight time among these participants as the cutoff (≥ 90 flight hours) for selecting the baseline group. These 22 baseline participants met or exceeded this cutoff and were included in the analysis. The tailored interface group had fewer participants than the baseline group. This difference may partly reflect the timing of data collection near the end of the semester, when fewer students were available to participate. In addition, 12 of 23 individuals screened met the instrument approach experience eligibility requirement. Not all participants were used for every analysis. For the subjective analysis, I excluded two baseline participants who had logged more than 1000 hours, because subjective workload and situation awareness ratings are anchored to expectations shaped by flight experience, and these two participants' experience levels were substantially higher than the rest of the sample. This produced a baseline subjective group of n = 20 (mean 235.03 flight hours, SD 115.30). I retained all 22 baseline participants for the physiological, flight performance, and gaze analysis because these measures reflect real-time responses that are less susceptible to experience-based self-report anchoring. In the tailored interface group, one participant had previously completed the baseline experiment and was excluded from all analyses, producing n = 11 (mean 200.66 flight hours, SD 93.12). For the physiological analysis, I excluded two additional tailored interface participants (n = 9): one session was unintentionally paused, allowing undue familiarity with the scenario, and one participant's sensor data could not be extracted.

Before data collection, the study targeted a sample size consistent with detecting a large interface effect in a two-group comparison. Using a two-sided $\alpha = 0.05$, 80% power, and an expected standardized effect size of approximately $d = 0.80$, the estimated requirement was approximately 25 participants per group. This target reflected the practical constraints of licensed pilot recruitment with instrument experience and simulator-based physiological data collection. However, detecting smaller effects would require larger samples; for example, a moderate effect of $d = 0.50$ would require approximately 64 participants per group. Therefore, the achieved sample size limits the ability to detect small-to-moderate effects.

6.2 Experimental Setup

Participants completed the task in a flight-simulation environment equipped with Brunner aircraft controls, a Varjo XR-4 headset (VR headset), a Polar H10 heart-rate sensor (worn underneath the shirt), and, in the tailored interface condition, a decision-support interface (masked by the VR headset). Figure 18 shows the participant setup, including the heart rate sensor, Varjo XR-4 headset, interface display, and simulator control.

Figure 18 Experiment setup showing the Varjo XR-4, Polar H10 heart-rate sensor, flight controls, simulator displays, and tailored interface display

Each participant used the Brunner controllers to manage aircraft attitude, airspeed, altitude, and lateral path. The Varjo XR-4 headset provided the visual simulation environment and recorded gaze and pupil-diameter data during the approach. The interface was masked in the VR environment in the tailored-interface group; i.e., the display remained visible during the task. The heart rate sensor recorded the cardiac activity throughout the task. These instruments enabled the study to capture physiological, visual, and flight-performance responses in a single operational scenario. The tailored-interface group viewed the interface during the approach. The interface displayed aircraft state, terrain awareness, flight path, and performance-related information. The interface did not control the aircraft or replace the pilot's judgment. Instead, it organized critical information so that participants could detect risk cues and monitor the aircraft relationship to the approach path.

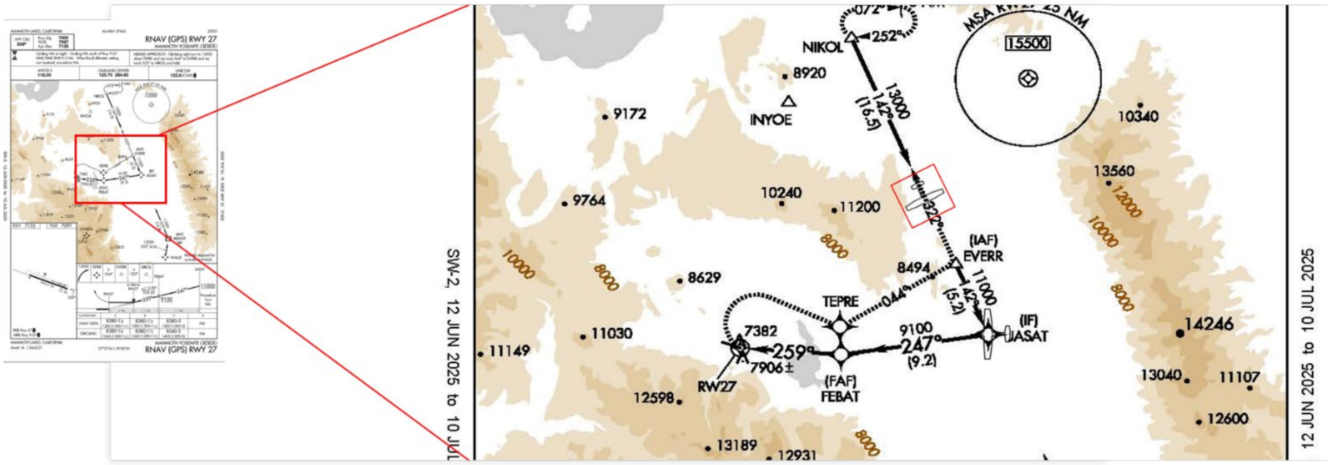

Figure 19 RNAV (GPS) RW 27 approach chart of Mammoth Yosemite Airport (KMMH).

The experiment used the Mammoth Yosemite RNAV RWY 27 approach as the flight scenario. Figure 19 shows the approach chart used to define the route, fixes, altitude constraints, and runway. We initialized the participant approximately 4.3 nautical miles from EVERR at 12,000 ft as shown in Figure 19. From that initial condition, participants flew toward EVERR, continued through JASAT and FEBAT, and proceeded toward Runway 27.

The scenario represented an IFR approach in mountainous terrain. The simulated weather conditions were 1.5 statute miles visibility with overcast cumulus clouds from 7,500 to 8,500 ft above mean sea level (MSL). These conditions limited outside visual references, forcing the participants to rely on information from their instruments, as they would in a real instrument approach. The mountainous terrain increased the importance of altitude control, lateral path tracking, and terrain awareness. The minimum descent altitude was 8,380 ft MSL, at a point approximately 4 NM from the runway 27 threshold. This altitude constraint created a critical decision point during the approach approximately 4 NM from Runway 27 threshold. Participants had to manage their descent profile while monitoring terrain clearance, lateral deviation, aircraft performance, and approach progress.

This scenario aligned with the dissertation's focus on CFIT risk. CFIT events often occur when pilots must control the aircraft attitude by clearing terrain proximity, maintaining altitude, and analyzing density altitude (aircraft performance at high altitude). The Yosemite RNAV RWY 27 approach required participants to integrate these demands while actively controlling the aircraft. Therefore, the scenario provided an appropriate test environment for evaluating whether the tailored interface supported pilot information processing during a CFIT-relevant task.

Each participant completed the same general procedure. First, we introduced the study and reviewed the consent information. After the participant agreed to continue with the experiment, they were taken to the relevant station and fitted with a heart rate sensor to collect resting heart rate data for 3 minutes. We then briefed the participant on the approach, fitted the Varjo XR-4 headset, checked the simulator and sensor setup, and started the scenario. The simulator placed the participant 4.3 nm from EVERR at 12000 ft. The participant then flew the Yosemite RNAV RW 27 approach. In both groups, data recording stopped if a participant landed, crashed, initiated a go-around, or decided to discontinue the experiment.

During the task, X-Plane 12 streamed aircraft state via various datarefs. Aircraft states refer to the simulator variables that describe the aircraft position, motion, and energy state during the task. The state variables used in this study are latitude, longitude and altitude. X-Plane exposes these variables as datarefs, which are named simulator data channels used to read aircraft and environment states. The datarefs used in this analysis are “sim/flightmodel/position/longitude”, “sim/flightmodel/position/latitude”, and sim/flightmodel/position/elevation. Similarly, the Varjo XR-4 streamed gaze and pupil diameter data, and the Polar H10 heart-rate sensor recorded cardiac activity. All these different streams were synchronized via Lab Streaming Layer (LSL), a lab recorder, as shown in Figure 20. After the scenario ended, participants completed a subjective questionnaire through Qualtrics. The questionnaire included NASA-TLX for perceived workload and SART for perceived situation awareness.

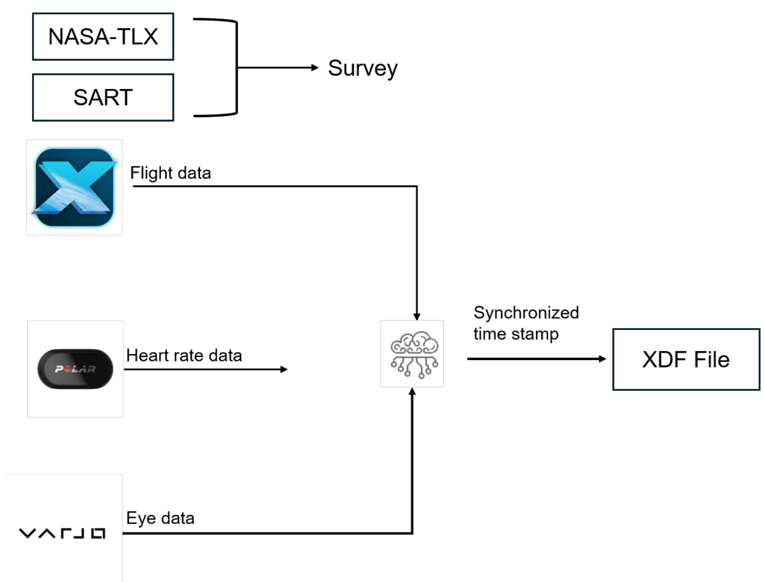

Figure 20 Synchronization of multi-source experimental data through Lab Streaming Layer into a single XDF file

The multimodal data structure allowed the study to evaluate the interface from several perspectives. Subjective measures captured participants' perceptions of workload and situation awareness after the task. Physiological measures captured task-related responses during the approach. Eye-tracking data captured visual attention and scanning behavior. Flight-path data captured operational performance during the approach.

6.3 Statistical Analysis for different measures

This study used nonparametric statistical tests because the groups were independent, unequal in size, and small in number. The Mann-Whitney U test compared the baseline and tailored interface groups without requiring the normality assumption of an independent-samples t-test (Fay & Proschan, 2010; Mann & Whitney, 1947). All statistical analyses were performed in Python. Data handling used the pandas (McKinney, 2010) and numpy libraries (Harris et al., 2020), and the Mann-Whitney U tests used the scipy.stats module in the scipy library (P. Virtanen et al., 2020).

I used Hedges' g (Hedges, 2008; Lakens, 2013) because it corrects standardized mean differences for small sample bias given by Equation 6

$$g = d \times j \qquad \text{Equation 6}$$

where $d = \left(\frac{\bar{X}_1 - \bar{X}_2}{S_p}\right)$, and $S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$ and $j = 1 - \frac{3}{4(n_1+n_2)-9}$. $\bar{X}_1$ and $\bar{X}_2$ represented group means, S_1 and S_2 represent group standard deviation, S_p represents the pooled standard deviation, and j represents the small-sample correction factor. I interpreted each result using p-values, descriptive statistics, and effect sizes together. A result with $p < 0.05$ indicated a statistically significant difference between groups. Non-significant results with a meaningful effect size indicated directional evidence that may warrant further testing with a larger sample.

The remaining sections present the results by outcome category. The subjective analysis 6.4 reports NASA-TLX workload and SART situation-awareness results. The physiological analysis 6.5 and 6.6 reports HRV and pupil-diameter results. The flight-path analysis reports lateral, vertical, and three-dimensional deviation results in section 6.7. The visual-attention analysis reports gaze behavior in section 6.8.

6.4 Subjective Analysis

I used the NASA-TLX (Hart & Staveland, 1988) to evaluate pilot workload across six dimensions, and SART (Taylor, 2017) to evaluate pilot situation awareness. The baseline group completed the approach without the tailored interface, whereas the interface group completed it with the tailored interface. Because different participants completed the baseline and interface conditions, I treated the comparison as a between-subjects analysis. The subjective analysis focused on two main questions: (1) Does the tailored interface reduce perceived workload, and (2) Does the tailored interface improve perceived situation awareness? I used descriptive statistics, non-parametric group comparisons, and effect sizes to interpret the results because of the small number of participants in the tailored interface group. After the flight task, participants completed NASA-TLX and SART questionnaires in Qualtrics. NASA-TLX is used to measure the perceived workload across its six dimensions mental demand, physical demand, temporal demand, performance, effort and frustration. SART is used to measure perceived situation awareness using demand, supply, and understanding dimensions. The questionnaires were available for participants in both interface conditions.

6.4.1 Data Preprocessing

The final subjective dataset included 20 baseline participants and 11 tailored-interface participants. The subjective data set contained the six NASA-TLX slider scores for *Mental Demand, Physical Demand, Temporal Demand, Performance, Effort,* and

Frustration. Each slider score ranged from 0 to 100. Higher values generally indicated a greater level of the corresponding dimension, except for the performance item, which used a positive direction. Specifically, the performance item represented perceived success or satisfaction, where higher values indicated better perceived performance. The pairwise comparison frequency counts were recorded for each NASA-TLX dimension. Each pairwise column contained two NASA-TLX dimensions, formatted as X: Y. A response of “1” indicated that the participant selected the first dimension (X in this case), and a response of “2” indicated that the participant selected the second dimension (Y in this case). For example, if the pairwise column label read “Mental Demand: Physical Demand” and the participants response was “1”, that response was added to the Mental Demand count by 1. If the participants response equaled “2”, I added one count to Physical Demand. After each computation, I verified that the total NASA-TLX pairwise count was 15 for each participant.

The analysis computed both raw and weighted NASA-TLX scores. Before computing these scores, the analysis reverse-coded the *Performance* item because the questionnaire coded *Performance* in the positive direction. In the questionnaire, higher performance scores indicated higher perceived success or satisfaction. However, NASA-TLX workload scoring requires that all dimensions point in the same direction, with higher scores indicating greater workload or poorer task experience. Therefore, the analysis computed a workload-direction performance score as given by Equation 7:

$$Performance_{TLX} = 100 - Performance_{Success} \qquad \text{Equation 7}$$

The original Performance score remained separate as a perceived performance/success outcome. The raw NASA-TLX score equaled the arithmetic mean of the six workload dimensions as shown in Equation 8.

$$NASA - TLX_{Raw} = \frac{MD + PD + TD + PR_{TLX} + EF + FR}{6} \qquad \text{Equation 8}$$

where MD is Mental Demand, PD is Physical Demand, TD is Temporal Demand, PR_{TLX} represents the reverse-coded performance workload score, EF represents Effort, and FR represents Frustration.

The weighted NASA-TLX score resulted from multiplying each dimension by its corresponding pairwise frequency count, summing across all six dimensions, and then dividing by 15 as shown in Equation 9

$$NASA - TLX_{Weighted} = \frac{(MD \times W_{MD}) + (PD \times W_{PD}) + (TD \times W_{TD}) + (PR_{TLX} \times W_{PR}) + (EF \times W_{EF}) + (FR \times W_{FR})}{15} \qquad \text{Equation 9}$$

SART evaluated pilots' perceived situation awareness after the approach task. The SART items represented three components of situation awareness: Demand, Supply, and Understanding. The Demand component included Instability, Complexity, and Variability. The Supply component included Arousal, Concentration, Division of Attention, and Spare Attentional Capacity. Lastly, the understanding component included Information Quantity, Familiarity, and Information Quality.

I computed SART demand, supply, understanding, and overall scores for each participant. The sum of instability, complexity, and variability represented demand as shown in Equation 10

$$SART_{demand} = Instability + Complexity + Variability \qquad \text{Equation 10}$$

Similarly, the sum of arousal, concentration, division of attention, and spare attentional capacity represented supply as shown in Equation 11

$$SART_{Supply} = Arousal + Concentration + Division + Spare \qquad \text{Equation 11}$$

Understanding equaled the sum of information quantity, familiarity, and information quality and is shown in Equation 12

$$SART_{Understanding} = Information\ Quantity + Familiarity + Information\ Quality \qquad \text{Equation 12}$$

The overall SART score is then calculated as shown in Equation 13

$$SART_{overall} = SART_{Understanding} - SART_{Demand} + SART_{Supply} \qquad \text{Equation 13}$$

Higher SART overall scores indicate better perceived situation awareness. Higher SART Understanding and Supply represented better outcomes. Lower SART Demand represented a better outcome because lower demand indicated a less demanding situation.

6.4.2 Statistical Analysis

The comparison of baseline and tailored-interface groups used Mann-Whitney U tests because the groups were independent and the interface sample size was small (Fay & Proschan, 2010; Mann & Whitney, 1947). Mean, standard deviation, and median described each measure’s group-level trends and variability. Hedges' g served as the effect-size estimate because it corrects Cohen’s d for small sample sizes (Hedges, 2008; Lakens, 2013). The analysis evaluated statistical significance at $\alpha = 0.05$. However, the interpretation of results used both p-values and effect sizes because the small sample size limited statistical power. Statistically nonsignificant results with moderate or large effect sizes provided preliminary evidence of directional effects rather than definitive evidence of improvement.

For workload-direction NASA-TLX measures, lower scores indicated better outcomes. These measures included raw NASA-TLX, weighted NASA-TLX, Mental Demand, Physical Demand, Temporal Demand, Performance-related workload, Effort, and Frustration. For the original Performance success scores, higher scores indicated better perceived performance. For SART, higher scores indicated better outcomes for Supply, Understanding, and Overall SART, while lower scores indicated better outcomes for Demand.

6.4.2.1 NASA-TLX Results

This study compared NASA-TLX scores between the baseline group and the tailored-interface group and computed both raw and weighted NASA-TLX workload scores after reverse-coding performance, as shown in Table 13. The original performance item measured perceived success, where higher values indicated better perceived task performance. Therefore, the NASA-TLX workload scoring used the reverse-coded performance score, while the original Performance Success score remained a separate outcome.

Table 12 Descriptive statistics (Mean, median, U and Hedges g) of NASA-TLX response given by participants from both baseline and tailored group

Measure	**Baseline Mean ±SD**	**Baseline Median**	**Interface Mean ±SD**	**Interface Median**	**Mean Difference**	**U**	***p***	**Hedges g**
NASA_TLX Raw	48.65 ± 14.54	49	42.69 ± 13.73	44.66	5.95	139	*0.23*	0.40
NASA_TLX Weighted	54.76 ± 17.15	58.33	48.33 ± 14.57	57.33	6.42	138	*0.26*	0.38
Mental Demand	74.75 ± 17.92	81	70.90 ± 14.52	70	3.8	127	*0.49*	0.22
Physical Demand	37.40 ± 25.83	30.50	35.09 ± 23.61	26	2.3	110	*1*	0.08
Temporal Demand	42.10 ± 20.53	33.5	41.36 ± 26.60	34	0.73	119.5	*0.70*	0.03
Performance (modified)	41.65 ± 30.80	36	22.63 ± 26.00	19	19.01	156	*0.06*	0.63
Effort	55.75 ± 25.43	64	54.90 ± 22.54	52	0.84	113	*0.91*	0.02
Frustration	40.25 ± 24.99	38	31.27 ± 34.77	16	8.97	142	*0.19*	0.31
Performance Original	58.35 ± 30.80	64	77.36 ± 26.00	81	-19.01	64	*0.06*	-0.65

The tailored interface group reported lower overall workload than the baseline group by 5.95 points. The baseline group had a raw NASA-TLX score of 48.65 ± 14.54, while the interface group had a raw score of 42.69 ± 13.73. The Mann-Whitney U test did not show a statistically significant difference ($U = 139$, $p = 0.23$). However, the effect size indicated a small to moderate, directional reduction in workload (g = 0.40).

The weighted NASA-TLX score showed a similar pattern. The baseline group had a weighted workload score of 54.76±17.15, while the interface group had a score of 48.33 ± 14.57. This represented a 6.42 point reduction in weighted workload for the interface group. The difference did not reach statistical significance (U = 138, p = 0.26), but the effect size again indicated a small-to-moderate directional reduction (g = 0.38). These results suggest that the tailored interface may have reduced perceived workload, although the small sample limited statistical power.

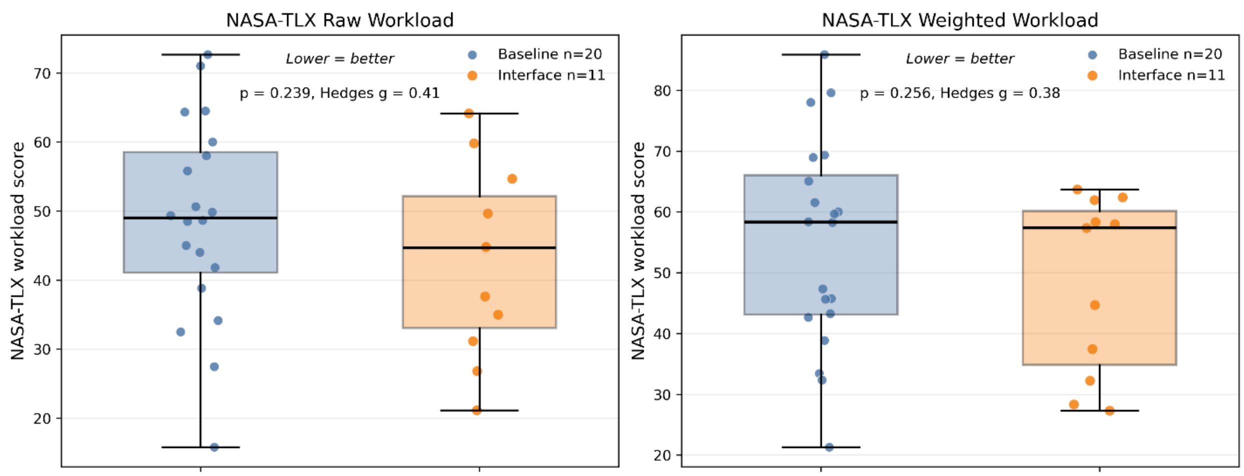

Figure 21 Raw and Weighted NASA-TLX workload scores for baseline and tailored interface groups

The NASA-TLX subscale results showed a more detailed pattern. The clearest difference was performance. The performance (modified for same directionality) item, the baseline group had a score of 41.65±30.80, whereas the tailored-interface group had a score of 22.63 ± 26.00. This difference was not statistically significant (U = 156, p = 0.06) and had a moderate-to-large effect size ($g = 0.63$). Because lower modified performance scores indicate lower perceived performance-related workload, this result suggests that participants using the tailored interface perceived better task performance.

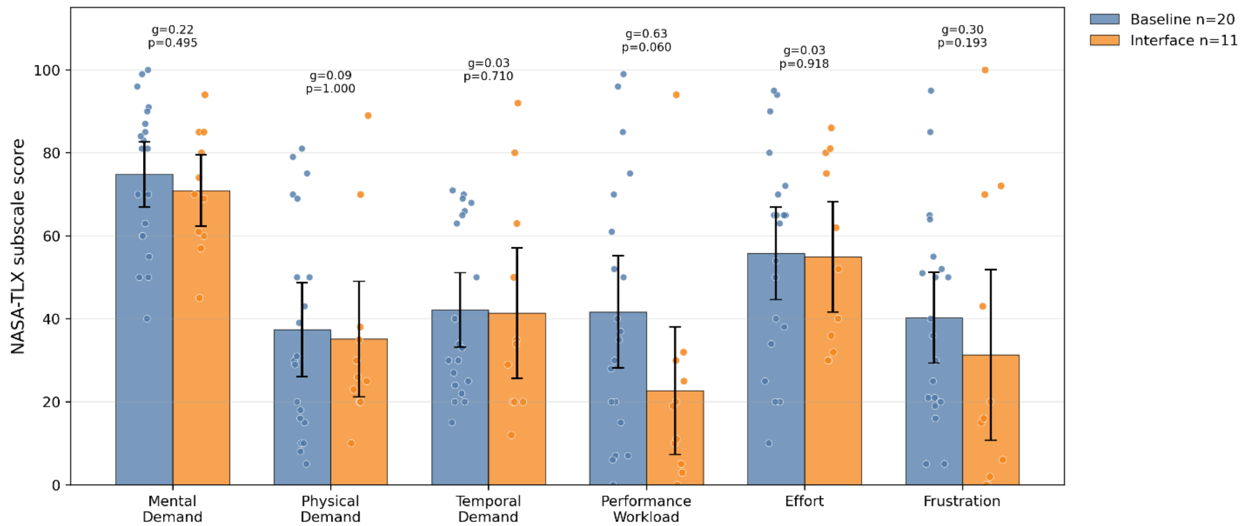

Figure 22 NASA TLX Scores of participants from baseline and tailored interface group.

Other subscales showed smaller or non-significant differences. Mental demand was slightly lower in the tailored-interface group (70.90 ± 14.52) than in the baseline group (74.75±17.92), but this difference was not statistically significant (U = 127, p = 0.49, g = 0.22). Physical demand was also similar between groups, with the baseline group at 37.40±25.83 and the tailored-interface group at 35.09 ± 23.61 (U = 110, p = 1, g = 0.08). Temporal demand was nearly identical across groups, with the baseline group at 42.10±20.53 and the tailored interface group at 41.36 ± 26.60 (U = 119.5, p = 0.70, g = 0.03). Effort was also nearly unchanged, with the baseline group at 55.75 ± 25.43 and the tailored-interface group at 54.90 ± 22.54 (U = 113, p = 0.91, g = 0.03). Frustration was lower in the tailored-interface group, but the difference did not reach statistical significance. The baseline group reported a frustration score of 40.25 ± 24.99, while the tailored-interface group reported 31.27 ± 34.77 which is an 8.97-point reduction in frustration for the tailored-interface group (U = 142, p = 0.19, g = 0.31)

Overall, the NASA-TLX results suggest that the tailored interface produced directionally favorable subjective outcomes. The tailored interface group reported lower raw and weighted workload, lower frustration, and significantly better performance success. However, the overall raw and weighted NASA-TLX workload scores did not reach statistical significance. Therefore, the NASA-TLX provide evidence that the tailored interface improved perceived task performance, with directional but non-conclusive evidence of reduced overall workload and frustration.

6.4.2.2 SART Results

I computed SART demand, supply, understanding, and overall scores for each participant. Lower demand scores represented better outcomes, and higher supply, understanding, and overall SART scores represented better outcomes.

The tailored interface group reported higher overall SART scores than the baseline group. The baseline group had an overall SART score of 22.78 ± 6.17, whereas the interface group had 25.35 ± 4.38. This represented a 2.56 point increase in overall perceived situation awareness for the interface group. The Mann-Whitney U test did not show a statistically significant difference ($U = 80$, $p = 0.36$), but the effect size indicated a moderate directional improvement ($g = -0.44$). The negative sign reflects the baseline minus interface effect-size direction; because higher SART overall scores indicate better situation awareness, this negative effect size favors the tailored-interface group.

Table 13 Descriptive statistics (Mean, median, U and Hedges g) of SART response given by participants from both baseline and tailored group

Measures (SART)	**Baseline Mean ±SD**	**Baseline Median**	**Interface Mean ±SD**	**Interface Median**	**Mean Difference**	**U**	***P***	**Hedges g**
Demand	10.89±3.22	10.65	12.22 ± 3.90	13.3	-1.33	87.5	*0.36*	-0.37

Supply	20.59±3.13	20.75	21.57 ± 3.02	21.8	-0.97	89.5	*0.40*	-0.30
Understanding	13.08±3.64	13.55	16.00 ± 3.47	15.6	-2.92	59.5	*0.03*	-0.79
Overall	22.78 ± 6.17	22.3	25.35 ± 4.53	24.85	-2.56	80	*0.22*	-0.44

As shown in Table 13, the strongest SART related improvement appeared in the understanding component. The baseline group had an SART understanding score of 13.08 ± 3.64, while the interface group had 16.00 ± 3.47. This represented a 2.92-point increase in situational understanding for the interface group. The difference reached statistical significance (U = 59.5, p = 0.03), with the effect size moderate to large (g = -0.79). This result suggests that the tailored interface may have helped pilots better understand the approach situation.

SART supply also increased in the tailored-interface group. The baseline group had a supply score of 20.59 ± 3.13, while the interface group had 21.57 ± 3.02. This represented a 0.97 increase in perceived attentional supply for the interface group. The difference did not reach statistical significance (U = 89.5, p = 0.40), but the effect size indicated a small directional improvement (g = -0.30). This result suggests that pilots using the interface may have perceived slightly greater attentional resources or better distribution of attention across the situation.

SART demand increased slightly in the tailored interface group. The baseline group had a demand score of 10.89 ± 3.22, while the interface group had a demand score of 12.22 ± 3.90. Because lower demand indicates a better outcome, this result suggests that the interface may have introduced some additional perceived demand.

However, the difference did not reach statistical significance (U = 87.5, p = 0.36), and the effect size was small to moderate (g = - 0.37).

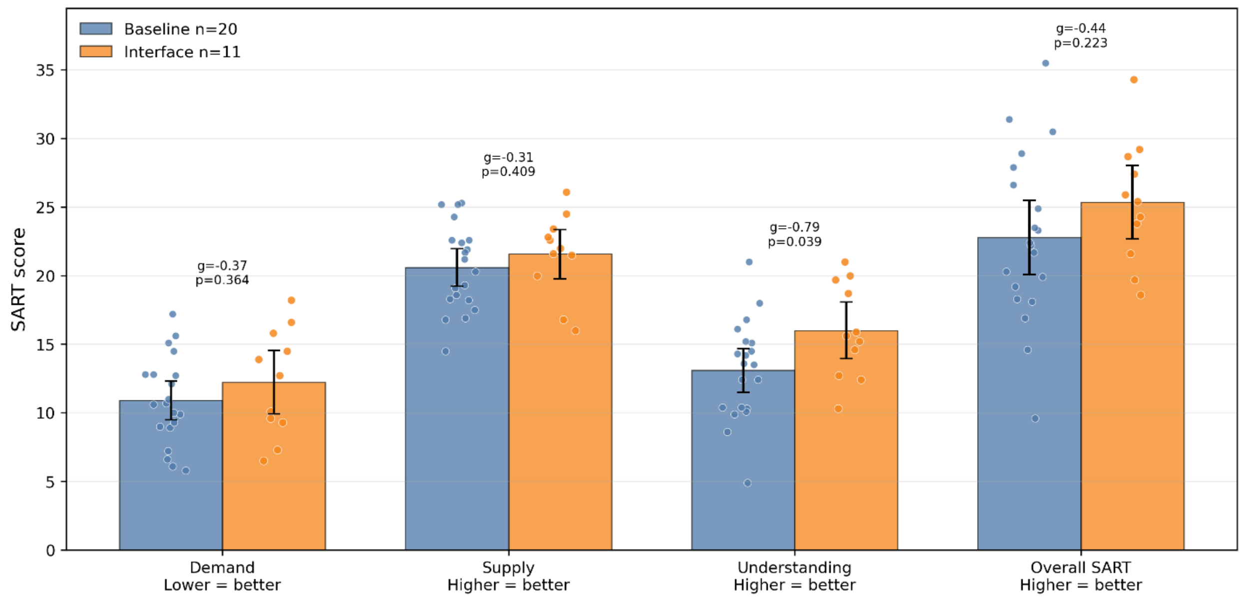

Figure 23 SART Demand, Supply, Understanding, and Overall SART scores for baseline and tailored-interface groups

Overall, the SART results suggest that the tailored interface improved perceived situational awareness, particularly situational understanding. The tailored interface group reported higher supply, understanding, and overall SART scores than the baseline group. The strongest effect was observed in SART understanding, as evidenced by statistical significance, which directly supports the purpose of the tailored interface: helping pilots interpret and understand relevant approach information. However, SART demand also increased slightly, suggesting that the interface may have required additional monitoring or information processing. Therefore, the SART findings suggest that the tailored interface improved situational understanding and overall situational awareness while potentially adding perceived demand.

6.4.3 Discussion

The subjective results provide evidence that the tailored interface improved pilot experience during the approach. As shown in Figure 24, the NASA-TLX results indicated lower raw and weighted workload for the interface group though these differences did not reach statistical significance. The clearest NASA TLX finding was for performance, as the tailored interface group reported significantly better perceived performance. The SART result strengthened this interpretation. The tailored interface group reported higher SART supply, understanding, and overall scores than the baseline group. The largest SART improvement occurred in the understanding component, suggesting that the interface supported pilots’ comprehension of the approach situation. This finding aligns with the tailored interface's goal of improving information processing and decision support during approach operations.

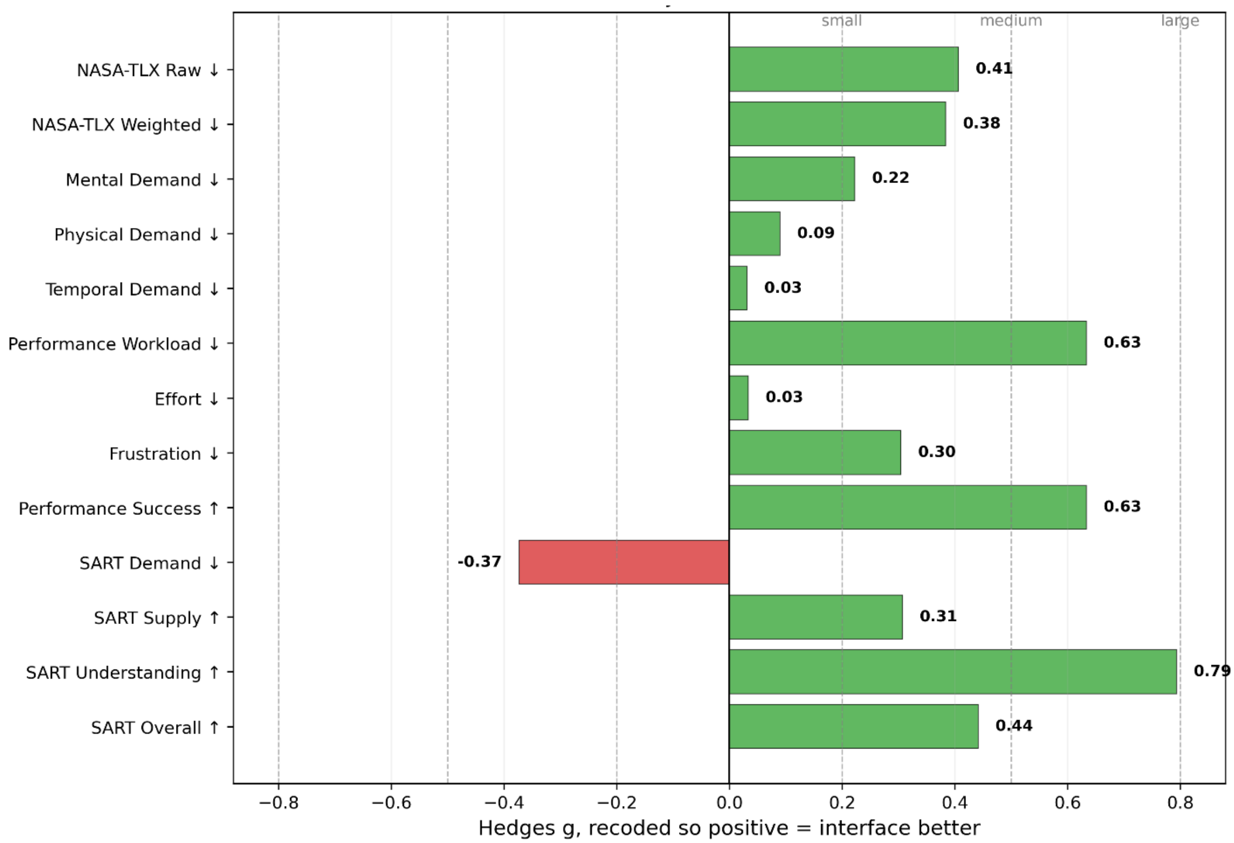

Figure 24 Effect-size summary for NASA-TLX and SART outcomes. Hedges' g values were directionally recoded so that positive values indicate better outcomes with the tailored interface

However, the subjective results also showed mixed effects. Temporal demand was similar between groups on the NASA-TLX, and SART demand increased slightly in the tailored interface groups. These findings suggest that the interface may have improved understanding of perceived performance while also introducing additional monitoring or increased information processing requirements. In other words, the interface may have helped pilots make sense of the situation, but it may not have fully reduced the demand associated with managing the approach.

The NASA-TLX and SART findings provide evidence of improved perceived performance and situational understanding, with directional but non-conclusive evidence of reduced overall workload. However, subjective ratings are post-task self-reports and may not fully capture pilots' real-time cognitive effort, visual attention, physiological responses, or flight-path performance during the approach. Therefore, the following section extends the analysis by examining pupil diameter, gaze behavior, heat maps, heart rate variability, and flight path deviation. The additional measures provide complementary evidence for determining whether subjective trends align with pilots' physiological workload, attention allocation, and flight performance outcomes.

6.5 Physiological Measures

This section presents the physiological analysis used to evaluate pilots' real-time workload response during the approach task. The subjective analysis in Section 6.2 showed the tailored interface produced directionally favorable changes in workload, perceived performance, frustration, and situation awareness. However, subjective measures rely on post-task self-report and may not fully capture the pilot's physiological response during task execution. Therefore, this study also analyzed physiological measures to examine whether the tailored interface influenced pilots’ autonomic and cognitive responses during the selected approach phase.

This study used heart-rate variability and pupil diameter as physiological indicators. Heart Rate variability provided information about autonomic regulation during the task, while pupil diameter provided information about visual-cognitive effort. The physiological analysis complemented the subjective NASA-TLX and SART results by evaluating whether pilots’ real-time responses aligned with their post-task ratings.

The first physiological analysis focused on heart rate and heart rate variability. This analysis examined whether a tailored interface changed pilots’ autonomic response during the selected approach phase compared with the baseline condition.

6.5.1 Heart Rate Variability

Heart Rate Variability (HRV) describes the variation in the time intervals between consecutive heartbeats. In workload related aviation research, reduced short-term HRV often reflects increased physiological engagement, attentional demand, or autonomic workload during task execution (Soares et al., 2025; Troyer et al., 2025). I used ln(RMSSD) for group comparisons because log-transforming the raw RMSSD values successfully reduced data skewness (Shaffer et al., 2020). RMSSD measures short-term variability in the intervals between consecutive heartbeats. The analysis compared the baseline group, which completed the approach without the tailored interface, with the tailored interface group, which completed the approach with the interface. Because different pilots contributed data to the two groups, this study treated the HRV comparison as a between-subject analysis.

6.5.1.1 Data Structure and Sample Size

I extracted heart rate and HRV metrics from the Electrocardiogram (ECG)-derived physiological dataset. The dataset included a resting segment (3 minutes) and a task segment (~ 17 minutes) for each participant. The resting segment represented the participants pre-task physiological state, and the task segment represented the participants physiological response during the selected approach phase.

The dataset included 22 baseline participants and 9 participants with a tailored interface. The HRV sample size differed from that of the subjective analysis because not all participants provided usable ECG data. I excluded records when the ECG signal did not support reliable HRV calculations, for one participant who took part in both experiments (therefore violating the between-subjects design), or when the experiment stopped immediately after the participant started the task.

6.5.1.2 Preprocessing

The analysis collected ECG data using a Polar H10 sensor, streamed via the Lab Streaming Layer (LSL), and saved in XDF format. The analysis processed the ECG recording separately for the resting and task segments. The resting segment represented the pre-task physiological baseline, and the task segment represented the selected approach phase used for the workload analysis.

The analysis first loaded each XDF file and searched for the polar H10 data stream. Some XDF files contained duplicate polar streams, so I selected the streams with the most samples. This step retained the stream with the most complete ECG recording and avoided empty or duplicate streams. For each valid Polar H10 ECG (heart rate sensor) stream, I extracted the raw ECG signal the sensor-reported heart rate, the reported RR values, and the LSL timestamps.

The Polar H10 transmitted ECG samples via Bluetooth bursts, causing multiple samples within a burst to share nearly identical LSL timestamps. To address this timing issue, the analysis reconstructed the ECG timestamps prior to resampling. The analysis detected burst boundary gaps larger than 0.005 seconds between consecutive LSL timestamps. Within each burst, the analysis reconstructed the sample times using the Polar H10 native ECG sampling rate of 130 Hz. This step restored the intended uniform temporal spacing of ECG samples within each burst.

After reconstructing the timestamps, I removed samples with missing ECG values then created a uniform time grid at 130 Hz and interpolated the ECG signal onto this grid using cubic interpolation. This step produced a uniformly sampled ECG signal for each recording. I also clipped extreme ECG amplitude values beyond 5σ from the mean to reduce the effects of large transient-amplitude artifacts before filtering and R-peak detection.

I filtered the resampled ECG signal using a fourth-order Butterworth bandpass filter with a passband from 5 to 15 Hz. This bandpass range is used to enhance the QRS complex and improve R-peak detection. The lower cutoff reduced baseline wander and slow movement-related drift, while the upper cutoff reduced high-frequency noise such as muscle activity.

I detected R-peaks from the filtered ECG signal using the *Neurokit2* , an open source Python package for neurophysiological signal processing (Makowski et al., 2021) Specifically I used *ecg_peaks* function, which identifies R-peak location from an ECG signal. The algorithm detects QRS complexes from the ECG signals and returns the sample locations of the R-peaks. Since the ECG signal had a sampling rate of 130 Hz, each detected R-peak corresponded to a sample location in the 130 Hz time series. I excluded any participants ECG recording that produced fewer than 10 detected R-peaks because that recording did not contain enough cardiac cycles for HRV calculations (Krause et al., 2023; Wehler et al., 2021).

After R-peak detection, RR intervals is computed from the difference between consecutive R-peak indices is given by Equation 14

$$RR_i = \frac{RR_{i+1} - R_i}{f_s} \times 1000 \qquad \text{Equation 14}$$

where R_i is the sample index of the i^{th} R-peak and $f_s = 130\ Hz$. This calculation produced RR intervals in milliseconds.

I then applied RR-interval artifact screening and correction. The screening removed RR intervals using lower and upper bounds of 300 ms and 1500 ms. These bounds correspond approximately to heart rates between 200 bpm and 40 bpm (Berntson et al., 1997; Tarvainen et al., 2014). If fewer than 10 valid RR intervals remained after this screening step, the analysis excluded these recordings from HRV calculations (Buysse, 2025).

For the remaining RR intervals, I computed the median RR interval and the median absolute deviation (MAD). It then identified potential artifacts using an adaptive threshold.

$$Threshold = \max(0.20 \times MedianRR, 3 \times MAD) \qquad \text{Equation 15}$$

Equation 15 marked an RR interval as an artifact if its absolute deviation from the median RR interval exceeded this threshold. This rule identified unusually short and long RR intervals that could reflect missed, extra, or ectopic beats, or detection errors.

When a participant ECG segment contained artifact intervals but retained sufficient non-artifact RR intervals, I corrected artifact intervals using cubic interpolation from the surrounding non-artifact intervals. Various literature recommends interpolation when RR-interval series contain artifacts, and cubic spline interpolation is a common approach (Alcantara et al., 2020; Peltola, 2012) After interpolation, it clipped the corrected RR intervals to the physiological range of 300-1500 ms. This correction preserved the continuity of the RR-intervals sequence while limiting the influence of non-physiological intervals on HRV metrics. This approach produced a corrected RR-interval series for each participant and segment. I then used this corrected RR series to compute heart rate and HRV metrics.

6.5.1.3 HRV Feature Extraction

I computed time-domain HRV metrics from the corrected RR intervals. I extracted RMSSD because RMSSD captures short-term beat-to-beat variability and serves as a useful indicator of vagal modulation during the task. RMSSD is then computed as shown in Equation 16

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(RR_{i+1} - RR_i)^2} \qquad \text{Equation 16}$$

where RR_i represents the i^{th} valid interbeat intervals and N represents the number of valid interbeat intervals in the segment.

Since RMSSD values can show right skewness (Kobayashi et al., 2012), I transformed RMSSD using the natural logarithm lnRMSSD (Plews et al., 2012). lnRMSSD is separately computed for the resting and the selected approach-task segment. I then computed the resting to task change as shown in Equation 17.

$$\Delta lnRMSSD = lnRMSSD_{Task} - lnRMSSD_{Resting} \qquad \text{Equation 17}$$

A negative ΔlnRMSSD indicates that HRV decreased from rest to task. A positive value indicates that HRV increased from rest to task.

Each participant contained two segments: a resting segment and a task segment. The resting segment lasted about 3 minutes and represented each participants pre-task physiological baseline. The task segment represented the selected approach phase and lasted approximately 15-17 minutes, depending on the participants recording and the availability of usable ECG.

I retained participants with valid resting and task ECG-derived HRV metrics. I excluded records when the ECG signal did not support reliable R-peak detection or HRV computation. The analysis also excluded records in which participants repeated the experiments and the repeated trial did not constitute an independent observation, or when the task stopped immediately after the participant began the experiments. After applying these criteria, the final HRV analysis included 22 baseline participants and 9 participants with a tailored interface.

Table 14 ECG/HRV preprocessing and sample size summary

Processing Item	Description
Sensor	Polar H10 ECG
Data Stream	LSL to XDF
Native ECG sampling rate	130 Hz
Input Files	Resting and Task XDF Files
Resting Segment	~ 3 minutes
Task Segment	~ 15-17 minutes
Timestamp Correction	Bursted timestamp reconstruction using 130 Hz spacing
ECG interpolation	Cubic interpolation to a uniform 130 Hz grid
ECG amplitude handling	Clipped values beyond 5 SD from the signal mean
R-peak detection	Neurokit2

RR physiological Range	300-1500 ms
RR artifact rule	Deviation from median RR greater than max 20
RR artifact correction	Cubic interpolation from non-artifact RR intervals
Primary HRV metric	lnRMSSD
Final baseline sample	n = 22
Final tailored interface sample	n = 9

6.5.1.4 Statistical Analysis

I analyzed the final HRV dataset so that each participant contributed one resting lnRMSSD value, one task lnRMSSD, and one ΔlnRMSSD value. The analysis retained resting and task heart rate, resting and task RMSSD, and resting-to-task changes for descriptive reporting. In addition, I used Wilcoxon signed-rank tests to examine within-subject changes from rest to task, separately within each group. I used the Mann-Whitney U test to compare the baseline and tailored interface groups because the two groups were independent, and the tailored interface HRV sample was small. The analysis reported descriptive statistics, including the mean and standard deviation. For between-group comparisons, it reported the Mann-Whitney U statistic, p-value, and rank-biserial effect size. I interpreted the p-value alongside the effect sizes because the tailored interface group contained only 9 valid ECG records.

Table 15 Heart Rate and HRV comparison between baseline and tailored interface groups

Measure	**Baseline Mean ±SD**	**Tailored Mean ±SD**	**Test/Statistic**	***p***	**Effect Estimate (r)**
Resting lnRMSSD	4.60±0.36	4.18±0.90	Mann-Whitney	0.19	-0.30
Task lnRMSSD	4.63±0.32	3.86±0.66	Mann-Whitney	0.02	-0.73

ΔlnRMSSD	0.03±0.18	-0.32±0.64	Mann-Whitney	0.10	-0.38

The baseline and tailored interface groups showed similar mean heart rate values. Resting heart rate was 82.3±13.6 bpm in the baseline group and 79.3±8.3 bpm in the tailored interface group. Task heart rate was also similar between groups, with the baseline group showing 83.4±11.8 bpm and the tailored interface group showing 84.1±8.2 bpm. The between-group comparison of resting-to-task heart rate change was not statistically significant (p=0.231). These results suggest that the mean heart rate did not differentiate the baseline and tailored interface conditions. However, the HRV results showed a different pattern. The baseline group showed very little change in RMSSD from resting to task, with resting RMSSD of 108.8±69.5 ms and task RMSSD of 108.7±48.6 ms. In contrast, the tailored interface group showed a decrease in RMSSD from 92.9±81.5 ms during rest to 56.7±31.6 ms during the task. Mean heart rate and RMSSD describe different cardiac responses. The mean heart rate gives the average number of beats per minute, whereas RMSSD reflects short-term beat to beat variability and parasympathetic modulation (Shaffer & Ginsberg, 2017). Therefore, the tailored interface group did not show a faster average heart rate, but it showed less beat-to-beat variability during the task. This pattern suggests a reduction in parasympathetic modulation, which may reflect greater attentional regulation or cognitive effort during the tailored-interface condition (Shaffer & Ginsberg, 2017). The lnRMSSD results showed the clearest group difference. The baseline group had a resting lnRMSSD of 4.60± 0.36 and a task lnRMSSD of 4.63±0.32. In contrast, the tailored interface group had a resting lnRMSSD of 4.18±0.90 and a task lnRMSSD of 3.86±0.66. The between-group comparison for task lnRMSSD was statistically significant (p = 0.02), with a large effect size (r = -0.73). The resting-to-task change in lnRMSSD further showed a directional pattern. The baseline group showed almost no change from rest to task, with ΔlnRMSSD 0.03 ± 0.18, whereas the tailored interface group showed a reduction in ΔlnRMSSD = -0.32 ± 0.64. However, the between-group comparison of ΔlnRMSSD did not reach conventional statistical significance (p=0.10), although the effect estimate was moderate (r= -0.38). The tailored interface group showed less natural variation between heartbeats during the task than the baseline group. Their heart beats were not necessarily faster, but the timing between beats became more regular. Since lnRMSSD reflects reduced short-term heart rate variability, this pattern could indicate that participants in the tailored group required more attention.

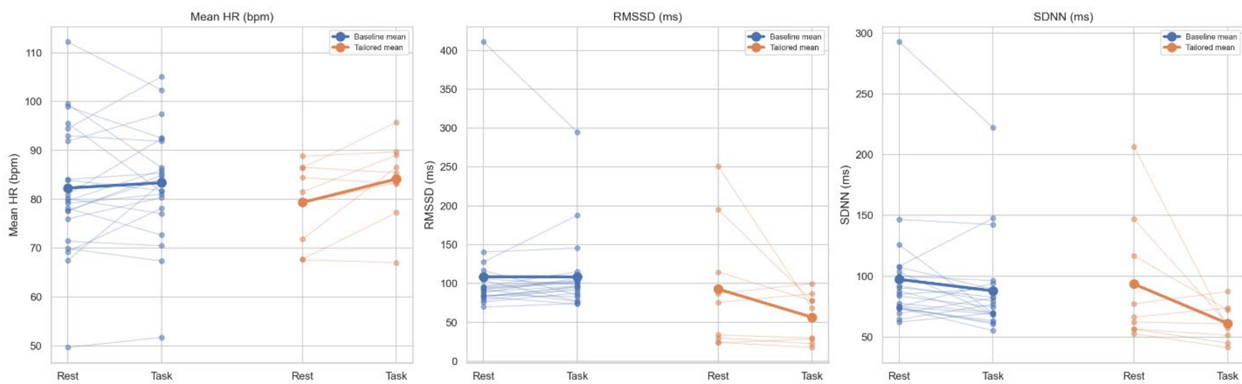

Figure 25 Resting-to-task changes in mean heart rate, RMSSD, and SDNN for the baseline and tailored interface groups. Thin lines represent individual participants, and thick lines represent group means. (A) mean heart rate, (B) raw RMSSD, and (C) SDNN

However, the rest-to-task change did not reach statistical significance, so this result should not be interpreted as confirmed difference in how much each group changed from rest to task.

6.5.2 Discussion

The HRV analysis adds physiological context to the subjective findings. The subjective analysis results suggested that the tailored interface reduced frustration, improved perceived performance, and improved situation understanding. The HRV results showed that task lnRMSSD was significantly lower in the tailored interface group than in the baseline group, while the resting-to-task change in lnRMSSD showed a directional but non-significant reduction. This pattern suggests a stronger task-related reduction in vagal modulation during the approach.

This result does not necessarily indicate that the tailored interface increased stress. A lower lnRMSSD can reflect greater physiological engagement, attentional demand, task involvement, or a workload-related autonomic response. In this study, the tailored interface group also reported lower frustration and higher situational understanding. Therefore, the HRV results suggest that pilots may have actively engaged with the tailored interface during the approach, yet still perceived it as useful. Mean heart rate did not show meaningful differences, indicating heart rate alone did not capture the more subtle physiological difference observed in lnRMSSD. This result supports the use of HRV, rather than heart rate alone, to evaluate workload-related physiological responses.

The tailored interface sample remained smaller than the baseline sample, which limits the strength of HRV interpretation. Therefore, these findings should be interpreted as evidence of a task-related physiological difference, with the change-score results treated as directional. Overall, the HRV results suggest that the tailored interface may have improved subjective understanding and perceived performance while still requiring active physiological engagement during task execution. This finding creates an important bridge to the next physiological measure. HRV summarized the autonomic response over the selected approach, but it did not identify moment-to-moment changes in visual cognitive effort. Therefore, the next section examines pupil diameter to assess whether a tailored interface affects cognitive effort during specific phases of the approach.

6.5.3 Pupil Dilation

This study used pupil diameter as a physiological measure of visual-cognitive effort during the selected approach phase. Pupil diameter can change in response to cognitive effort, attentional demand, arousal, and information processing load (van der Wel & van Steenbergen, 2018). Therefore, pupil diameter provides a useful, complementary measure for evaluating whether the tailored interface altered pilots' real-time cognitive responses during the task. However, pupil diameter does not reflect cognitive workload alone. It can also change because of luminance, display brightness, gaze direction, accommodation, headset fit, eye-tracker quality, iris characteristics, baseline arousal, and individual physiological differences. For this reason, I did not interpret raw pupil diameter as a standalone measure of workload. Instead, this study used pupil diameter as one component of a larger triangulated analysis that also included NASA-TLX, SART, HRV, gaze behavior, and flight-path deviation. The pupil analysis addressed the question "Did the tailored interface produce a measurable difference in pupil-based visual-cognitive effort compared with the baseline condition during the approach?"

To answer this question, the analysis extracted task-period pupil data from the Varjo XR-4 eye-tracking stream, cleaned the pupil signal, corrected the sampling-rate mismatch between groups, computed participant-level pupil metrics, and compared the baseline and tailored-interface groups using between-subject statistical tests.

6.5.3.1 Data Structure

The analysis extracted pupil data from the VARJO XR-4 eye-tracking data recorded through LSL and stored in XDF files.

The analysis used only task-period files because the dataset lacked a separate resting-state pupil recording. Therefore, the pupil analysis compared pupil behavior during the selected task period rather than comparing resting and task pupil values. The final pupil dataset included 31 accepted task files: baseline n = 22 and tailored n = 9. All 31 files met the final data-quality requirements after preprocessing. The baseline files had an original sampling rate of approximately 4 Hz, while the tailored interface files had an original sampling rate of approximately 100 Hz. To address the difference in sampling rate, I resampled both groups to a common 4 Hz analysis rate before computing final pupil metrics. The corrected pupil pipeline explicitly implemented sampling-rate correction, blink interpolation for short gaps, linear detrending, median-based pseudo-baseline normalization, and per-participant quality reporting.

Table 16 Pupil-diameter preprocessing and data-quality criteria

Processing Item	**Decision used in this study**
Sensor	VARJO XR-4
Data Stream	LSL to XDF
Input files	XDF through LSL
Resting pupil recording	Not available
Pseudo-Baseline	First 60s of selected task period (initial-task reference)
Baseline original sampling rate	Approximately 4 Hz

Tailored original sampling rate	Approximately 100 Hz
Resampled sampling rate	4 Hz
Eye-openness threshold	≥ 0.5
Physiological pupil range	1.5-8.0 mm
Short blink interpolation	Gaps < 300 ms
Outlier detection	Median Absolute Deviation (3 MAD)
Minimum clean binocular data	70%
Minimum task duration	300 s
Primary Statistical Unit	Participant-level metrics

6.5.3.2 Preprocessing

Raw pupil diameter varies substantially across participants. One pilot may naturally have a larger or smaller pupil diameter because of individual physiology, iris characteristics, age, arousal, fatigue, display viewing behavior, or environmental lighting. Therefore, a direct comparison such as “the tailored group had a pupil diameter of 2.54 mm and the baseline group had a pupil diameter of 2.57 mm” does not, by itself, isolate the effect of the interface.

I therefore treated the raw mean pupil diameter as a descriptive measure rather than the primary workload outcome. The main pupil outcomes were pseudo-baseline normalized measures. These measures compare each participants later-task pupil response to their initial-task pupil level. This within-participant normalization reduced the influence of between-person differences in raw pupil size. Because the task began immediately after the Varjo XR-4 recording, this study lacked a true resting pupil baseline. Therefore, I used the first 60s of the selected task period as an initial pseudo-baseline. This approach did not estimate true rest-to-task pupil dilation. Instead, it estimated how pupil diameter changed after the initial task reference period. This is important because the 60s already include task demand, visual stimulation, headset adaptation, luminance effect, and interface exposure. Therefore, this study interpreted pseudo-baseline-normalized pupil results as task-relative pupil changes rather than true resting-baseline changes. The original pupil recordings had a major sampling-rate mismatch: The sampling rate is calculated as shown in Equation 18

$$f_s = \frac{N_{samples}}{T_{end} - T_{start}} \quad \text{Equation 18}$$

where $N_{samples}$ is the number of eye-tracking samples and $T_{end} - T_{start}$ is the recording duration in seconds. This mismatch could bias pupil variability metrics. For example, a 100 Hz signal contains many more samples per second than a 4 Hz signal. If the analysis computed standard deviation, coefficient of variation, or windowed pupil metrics directly from the original samples, the tailored interface group would contribute much denser data than the baseline group. This would make the two groups less comparable.

To correct this issue, the analysis resampled all cleaned (discussed later) binocular pupil signals to a common 4 Hz analysis rate. I down sampled frequency to 4 Hz because the baseline group was recorded at approximately 4 Hz. This decision avoided artificially increasing the temporal resolution of the baseline data and placed both groups on the same effective time scale. After time binning, each participant contributed pupil data at the same analysis rate regardless of the original recording rate. The analysis used these 4 Hz resampled signals for all final pupil metrics and statistical comparisons.

I also retained the original sampling rate as a data-quality variable because the baseline and tailored-interface recordings used different original sampling rates.

Samples that had missing, non-numeric, zero, or negative values were discarded as invalid physiological pupil diameters. This validity check is performed for both the left and right eyes. The stream also included eye-openness values for each eye. Eye openness was represented on a 0-1 scale, with higher values indicating greater eye openness. I retained pupil samples only when the corresponding eye-openness was at least 0.5.

This threshold is important because pupil estimates become unreliable when the eye is partially closed or when the eyelid occludes the pupil. This filtering step removed samples associated with blinks, partial closure, and unreliable pupil estimation.

After applying the validity and openness filters, the analysis applied a physiological pupil-diameter range filter. It retained only pupil values between 1.5 mm ≤ Pupil diameter ≤ 8.0 mm (Spector, 1990). This range removed values that were too small or large to be plausible in the experimental context. Very small or large values likely reflect tracking errors, spike artifacts, segmentation errors, or invalid samples. After the filters, the short, blink-related gaps lasting less than 300 ms were interpolated using cubic interpolation. This process is important because prolonged missing periods may indicate sustained tracking loss, prolonged eye closure, or poor headset fit. Interpolating long gaps would create artificial pupil data and could bias the analysis. The blink interpolation step preserved short blink-related gaps while avoiding artificial reconstruction of longer periods of missing data. Pupil signals can show slow drift over time. This is because of gradual headset movement or changes in visual behavior. If the analysis does not account for drift, slow non-cognitive change may influence summary pupil metrics. Therefore, I removed linear drift from each pupil signal after blink interpolation. To achieve this, I fitted a linear trend to the valid pupil samples and removed the slope component. Nevertheless, the analysis preserved the participants' central pupil diameter, ensuring that the processed pupil signal remained interpretable in millimeters. This step removed slow linear drift while retaining participants' overall pupil-size levels.

After linear drift correction, I applied a median absolute deviation (MAD) outlier filter. Pupil data often contain spike artifacts and non-normal distributions, so applying this method is preferable to standard deviation-based filters, which are strongly affected by extreme values.

For each participant, the cleaned pupil signal is given by Equation 19

$$P = \{P_1, P_2, P_3, \ldots., P_N\} \qquad \text{Equation 19}$$

where P_i represents the i^{th} cleaned pupil-diameter sample, and N represents the total number of valid samples in the signal. The median of P_i is computed

This value represented the central pupil diameter in the participants' cleaned signals. Then the median absolute deviation is computed as shown in Equation 20

$$MAD = Median\ (|P_i - Median(P)|) \qquad \text{Equation 20}$$

I removed samples that exceeded the 3-MAD threshold. After this, I used the binocular average for the main pupil analysis because it provides a more stable pupil estimate than either eye alone, which is given by Equation 21

$$Pupil_{Avg}(t_i) = \frac{Pupil_{left}(t_i) + Pupil_{right}(t_i)}{2} \qquad \text{Equation 21}$$

where $Pupil_{Left}(t_i)$ and $Pupil_{Right}(t_i)$ represent the cleaned left and right pupil diameter at time t_i. The resulting binocular-averaged signal served as the primary pupil-diameter signal for each participant. The analysis required each accepted file to meet two inclusion criteria. The data should at least have 70% of the clean binocular data and the recording should at least last 300s.

These criteria ensured that each participant contributed sufficient pupil data for participant-level analysis. Altogether 31 files were accepted because they met the final data-quality requirements. The final clean binocular data percentages were high in both groups: the baseline group had 95.3%, and tailored group had 96.0% of the clean binocular data. These values indicated that the accepted pupil data had sufficient quality for group-level analysis.

As mentioned above, raw pupil diameter varies across individuals, so I used pseudo-baseline normalization to reduce between-person differences. Because no true resting-pupil recording was available, the analysis used the first 60s of the selected task period as the initial task reference period. Therefore, it is referred to as a pseudo-baseline rather than a true resting baseline. This distinction is important because the task had already started during this period, i.e., it included early task demands, interface exposure, headset adaptation, luminance effects, and visual scanning. Therefore, the pseudo-baseline does not estimate resting pupil diameter. This approach allowed me to relate each participants later-task pupil response to their own early-task pupil level.

Therefore, the normalized pupil outcomes represent task-relative changes rather than rest-to-task changes.

For each participant, I divided the resampled pupil signals into two parts $P_{BL} = \{P_i: t_i < 60s\}$ $P_{Task} = \{P_i: t_i > 60s\}$ where P_{BL} represents the first 60s pseudo-baseline pupil samples and P_{Task} represents the remaining task period samples. The median pupil diameter during the first 60 seconds is given by $Pupil_{PseudoBaselineMedian}$, and for the rest of the period, the median is calculated as $Pupil_{TaskMedian} = Median\ (P_{Task})$.

Now the z-scores are calculated as shown in Equation 22

$$z_{pupil} = \frac{Pupil_{TaskMedian} - Pupil_{PseudoBaselineMedian}}{MAD_{PseudoBaseline}} \qquad \text{Equation 22}$$

where $MAD_{PseudoBaseline}$ represents the median absolute deviation of the first 60s pseudo-baseline pupil samples. Then the percentage change from the pseudo-baseline is given by Equation 23

$$\%\Delta Pupil = \frac{(Pupil_{TaskMedian} - Pupil_{PseudoBaselineMedian})}{Pupil_{PseudoBaselineMedian}} \times 100 \qquad \text{Equation 23}$$

Positive values indicated that pupil diameter increased after the initial task-reference period. Negative values indicated that pupil diameter decreased after the initial task-referenced period. In summary, I computed six participant-level pupil outcomes from the cleaned and resampled binocular pupil signals, whose summary is presented in Table 17.

Table 17 Interpretation of each metric used in pupil dilation

Metric	Purpose	Interpretation
Mean Pupil Diameter	Descriptive Pupil Size	Raw group context, not primary workload inference

Pupil Standard Deviation	Pupil Variability	Larger values indicate greater variability
Pupil Coefficient of Variation	Variability normalized by mean	Helps compare variability relative to pupil size
Pupil Interquartile range	Robust Spread	Less Sensitive to Outliers than the standard deviation
Pseudo-baseline z-score	Task-relative pupil change	Positive values indicate larger later-task pupil diameter
Percent change from pseudo-baseline	Task Relative percent change	Positive values indicate an increase from the initial task reference.

6.5.3.3 Statistical Analysis

I used participant-level values for all statistical comparisons. Each participant contributed one value per pupil outcome. This choice avoided treating repeated pupil samples from the same participants as independent observations. I compared the baseline and tailored-interface groups using the Mann-Whitney U Test because the groups were independent and the tailored-interface sample was small. The analysis reported Mann-Whitney U Statistics, p-value, and the rank biserial effect size. Pupil analysis can sometimes reflect both workload-related and non-workload-related factors; therefore, results are reported with caution. Mainly, I emphasize pseudo-baseline-normalized outcomes because they account for each participants initial task-reference pupil level.

Table 18 Pupil diameter statistical comparison between baseline and tailored-interface group

Measure	**Baseline Mean ±SD**	**Tailored Mean ±SD**	**Mann-Whitney U**	**p-value**	**Rank-biserial r**
Mean pupil diameter, mm	2.57 ± 0.40	3.70 ± 1.46	64	0.13	0.35
Pupil SD, mm	0.15 ± 0.04	0.22 ± 0.02	51	0.03	0.48

Pupil CV	0.06 ± 0.01	0.06 ± 0.01	98	0.98	0.01
Pupil IQR,mm	0.20± 0.05	0.28 ± 0.11	50	0.03	0.49
Pseudo-baseline z-score	-0.35 ± 1.47	-0.26 ± 0.86	92	0.77	0.07
Percent change from pseudo-baseline, %	-1.08 ± 4.38	-1.10 ± 2.97	97	0.94	0.02

As shown in Table 18, the tailored interface group showed a numerically larger mean pupil diameter than the baseline group, although this difference did not reach statistical significance. The baseline group had a mean pupil diameter of 2.57±0.40 mm, whereas the tailored interface group had a mean pupil diameter of 3.70 ±1.46 mm. The Mann-Whitney U test did not show a statistically significant difference, $U = 64$, $p = 0.13$, $r = 0.35$. Because raw pupil diameter can vary across individuals for reasons unrelated to workload, this result served mainly as a descriptive comparison.

Pupil variability showed a clear group difference in the absolute metrics. The baseline group had a pupil standard deviation of 0.15 ± 0.04 mm. The tailored interface group had a pupil standard deviation of 0.22 ± 0.02 mm. This difference is statistically significant ($U = 51$, $p = 0.03$, $r = 0.48$). A similar pattern is observed for pupil IQR. The baseline group had a pupil IQR of 0.20±0.05 mm, while the tailored interface group had 0.28±0.11 mm. This comparison was also statistically significant ($U = 50$, $p = 0.03$, $r = 0.49$). These results suggest greater absolute pupil variability in the tailored interface group.

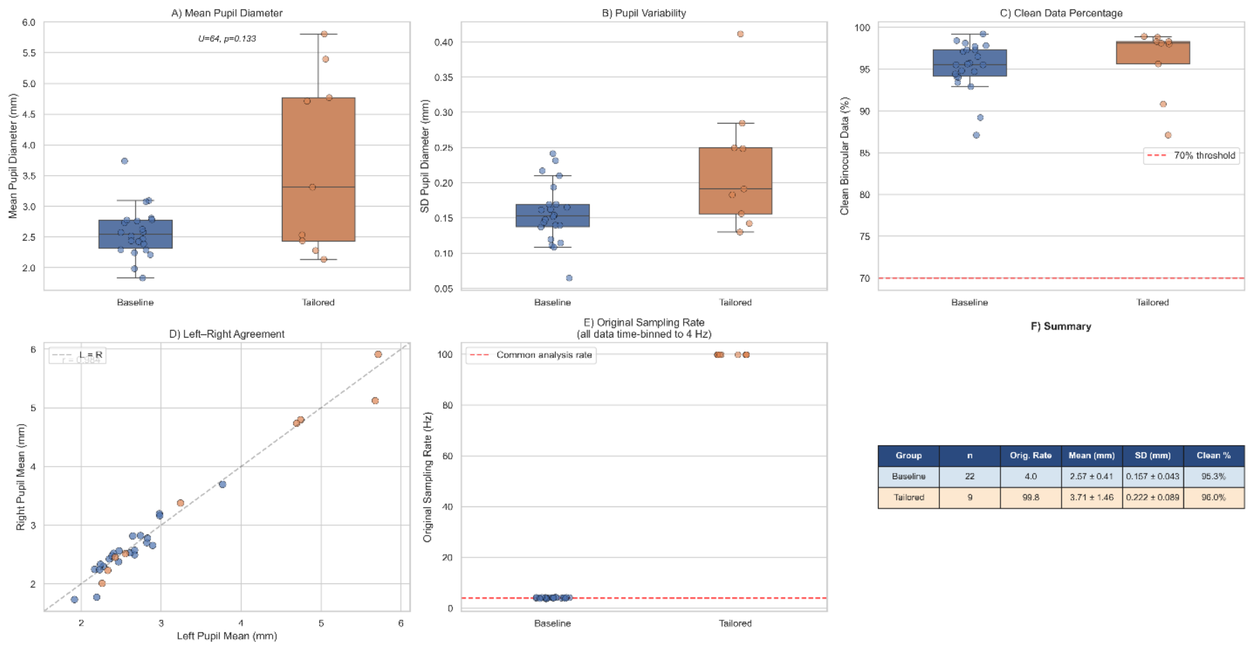

Figure 26 Pupil diameter data quality and group summaries for the baseline and tailored interface group

Figure 26 shows the mean pupil diameter (A), pupil variability (B), the percentage of clean binocular data (C), left-right agreement (D), the original sampling rate (E), and group-level summary statistics (F).

However, the coefficient of variation did not differ between groups. Both groups had a pupil CV of 0.06 ± 0.01, and the group comparison was not statistically significant ($U = 98$, $p = 0.98$, $r = 0.01$). This result suggests that although absolute pupil variability was higher in the tailored-interface group, relative variability scaled by mean pupil diameter was similar across groups.

The pseudo-baseline-normalized metrics also showed no clear group difference. The baseline had a pseudo-baseline z-score of -0.35±1.47, while tailored had -0.26±0.86. The group comparison was not statistically significant ($U = 92$, $p = 0.77$, $r = 0.07$). The percent change metric showed the same pattern. The baseline group changed by -1.08 ± 4.38% relative to the initial task-reference period. The tailored interface group changed by -1.10 ± 2.97%; again, the comparison was not statistically significant (U = 97, p = 0.94, r = 0.02). These results indicate that the tailored interface did not produce a measurable increase or decrease in pupil dilation relative to the initial task-reference period.

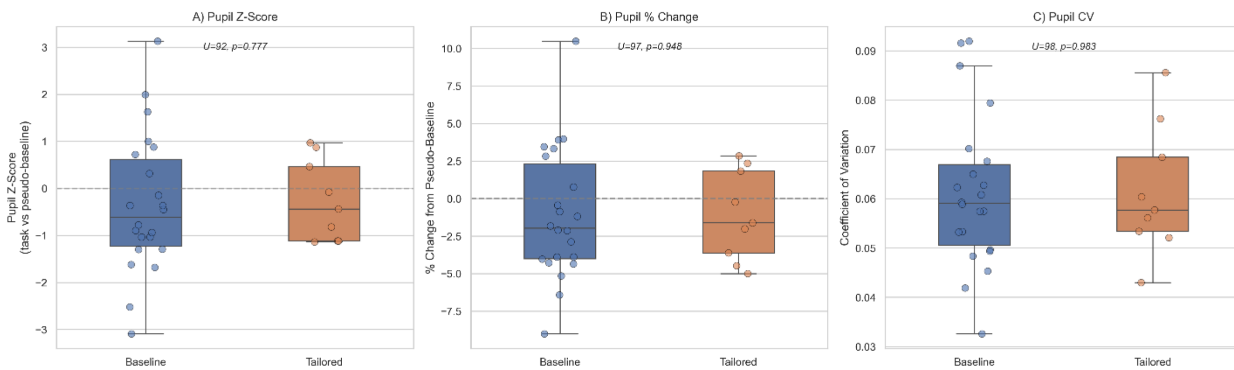

Figure 27 Pseudo-baseline normalized pupil dilation outcomes for baseline and tailored-interface groups

The first 60s of the task served as an initial-task pseudo-baseline because no resting pupil recording was available. Figure 27 shows pupil z-score (A), percent change from pseudo-baseline (B), and pupil CV (C). Values above zero indicate increased pupil diameter relative to the initial task reference period.

Overall, the pupil-diameter results were mixed. Absolute pupil variability, measured by pupil SD and IQR, was higher in the tailored-interface group. However, mean pupil diameter, pupil CV, pseudo-baseline z-score, and percent change from pseudo-baseline did not show statistically significant differences between groups. Therefore, the pupil results provide limited evidence of increased absolute pupil variability in the tailored-interface group, but they do not provide strong evidence of a normalized increase in pupil-based visual-cognitive effort.

6.5.3.4 Discussion

The pupil-diameter analysis provides a more nuanced physiological result than a simple null finding. The subjective analysis suggested that the tailored interface led to perceived performance improvements, reduced frustration, and improved situational understanding. The HRV analysis suggested a greater physiological engagement in the tailored-interface group. The pupil analysis showed higher absolute pupil variability in the tailored-interface group, as reflected by significantly higher pupil SD and IQR. However, normalized pupil outcomes, including pupil CV, pseudo-baseline z-score, and percent change from pseudo-baseline, did not differ significantly between groups. Pupil diameter is sensitive to cognitive effort, but it is also affected by non-cognitive factors such as display luminance, gaze direction, accommodation, arousal, headset fit, and individual physiology. Therefore, the significant differences in absolute pupil variability may reflect greater visual or physiological fluctuation during the task, but they cannot be interpreted on their own as conclusive evidence of increased cognitive workload.

The pseudo-baseline-normalized results are especially important. Raw pupil size and absolute variability can differ across individuals. The pseudo-baseline z-score and percent change measures compared each participants later-task pupil diameter with their own initial task reference period. These normalized measures also showed no significant difference. Therefore, the pupil analysis suggests that the tailored interface did not substantially increase or decrease pupil dilation after the initial task period.

There are several limitations to this analysis. First, this analysis did not collect a true resting pupil baseline. The first sixty pseudo-baseline periods occurred after task onset, so they may already include task demands, headset adaptation, display brightness effects, visual scanning, and interface exposure. Therefore, this analysis cannot estimate true rest-to-task dilation. It only estimates changes relative to each participants initial task period. Second, the original baseline and tailored-interface recordings had different sampling rates. I corrected this issue by time binning all signals to 4 Hz.

Overall, the pupil results do not contradict the subjective or HRV findings. Instead, they suggest that the tailored interface group showed greater absolute pupil variability without a corresponding normalized increase in pupil dilation. This may indicate that the tailored interface requires active visual engagement, but did not produce a strong group-level increase in pupil-based visual-cognitive effort after normalization.

However, subjective and physiological measures alone cannot determine whether the tailored interface improved operational task performance. NASA-TLX and SART describe pilots' perceived workload and situation awareness, while HRV and pupil diameter describe physiological response during the task. These measures do not directly show whether the pilot maintained the desired approach path more accurately. Therefore, the next section examines flight path deviation as an operational performance outcome. Specifically, I will analyze the lateral and vertical deviation to determine whether the tailored interface helped pilots maintain the desired approach path during the selected approach phase.

After evaluating flight-path deviation, this study examines visual attention behavior. The flight path deviation analysis indicates whether the aircraft stayed closer to the desired path, but it does not explain the pilot's distributed visual attention during the task. Therefore, the latter gaze-path and heat-map analysis examine where pilots looked during the approach and whether attention allocation helps explain the subjective, physiological, and flight performance results.

6.6 Flight Path Deviation Analysis

This section evaluates whether the tailored interface influenced pilots' operational flight-path performance during the selected approach phase. I previously examined subjective workload, situation awareness, HRV, and pupil diameter. These metrics describe the pilot’s perceived workload, situation awareness, physiological engagement, and visual cognitive effort. Yet, they do not directly show whether pilots maintained the intended aircraft flight path during the approach. So, this study analyzed flight path deviation as an operational performance measure.

The flight path deviation analysis focused on the RNAV approach path into Mammoth Yosemite Airport. I quantified path-following performance using lateral and vertical deviations, as well as the combined three-dimensional deviation. Lateral deviation measures how far the aircraft moved away from the ideal lateral approach path. Vertical deviation measured the extent to which the aircraft's altitude differed from the revised ideal vertical reference profile. The combined three-dimensional metric summarized lateral and vertical error in a single path-deviation measure. The analysis addressed the following question: *Did the tailored interface help pilots maintain the desired approach path more accurately than the baseline interface?*

Here, a lower deviation indicates better tracking of the desired approach path, while a higher deviation indicates poorer alignment with the reference path.

6.6.1 Data Source

The processing pipeline searched and extracted aircraft latitude, longitude, elevation, indicated airspeed, horizontal path, vertical path, pitch, bank, and ground speed when available from each participants' recorded file. Latitude, longitude, and elevation were required because these variables defined the aircraft's position in three-dimensional space. The code converted X-Plane elevation from meters to feet for altitude analysis.

The sample for this analysis $n_{Baseline} = 22$ and $n_{Tailored} = 9$. The baseline group included 18 go-around/missed approach decisions, 3 landings, and 1 crash. The tailored-interface group included 8 go-around/missed approach decisions and 1 landing. Since the scenario required a go-around decision, the missed approach outcome should not be treated as a failure. Instead, it represents the expected procedural decision when pilots did not have the required conditions to continue the approach. I retained the outcome labels for visualization and interpretation, but the main statistical analysis used continuous participant-level deviation metrics.

A reference path represented the desired Area Navigation (RNAV) centerline from the selected approach start point to Runway 27. I used this ideal path to compute cross-track error, along-track position, vertical reference altitude, path coverage, and distance-normalized deviation metrics. First, I loaded the X-Plane *earth_fix.dat* file and extracted waypoint names, latitudes, and longitudes. I stored these values in the database and verified that EVERR, JASAT, and FEBAT existed. Second, I loaded the X-Plane *apt.dat* airport database to extract the Runway 27 threshold location at KMMH. Here, I searched the KMMH airport record and parsed the runway definition line for runway designator 27. This step returned the latitude and longitude of the RW27 threshold, which I then added to the waypoint database as RW27. This analysis then created an artificial start point called REDX, which is not a standard published fix. REDX as a point is 7 nm from EVERR along a 322° bearing using the WGS84 ellipsoid. The final waypoint sequence used to build the ideal path was:

REDX → EVERR → JASAT →FEBAT → RW27

Each consecutive waypoint pair is then connected using WGS84 geodesic interpolation. Then, I generated points along each segment at approximately 75 m intervals and computed the cumulative along-track distance, s_m, from the beginning of the ideal path. The final ideal path file contained *idx, segment_id, wp_from, wp_to, lat, lon, and sm,* where *lat and lon* define the ideal lateral path*, segment_id* identifies the path segment*, wp_from and wp_to identify* the segment endpoints*,* and *sm* gives cumulative along-track distance in meters. This procedure made the lateral reference path reproducible by using X-Plane navigation data, the X-Plane runway threshold location, WGS84 geodesic calculations, and a fixed interpolation spacing, rather than a manually drawn path.

The following altitude values are used for the intended approach altitude constraints. REDX (Initialized point) = 12000 ft, EVERR = 13000 ft, JASAT = 11000 ft, FEBAT = 9100 ft, RW 27: 7097 ft.

In the scenario, pilots should have been placed near the appropriate early approach altitude. Nevertheless, the simulator scenario initialized participants at a lower altitude than intended, approximately 12,000 ft. In addition, linearly interpolating the published constraints yields a reference profile that rises from REDX (12,000 ft) to

EVERR (13,000 ft), then descends through JASAT and FEBAT. Because participants were initialized at approximately 12,000 ft, they could not match this early climb segment, which further inflated negative vertical deviation early in the approach for both groups. For these reasons, I treat vertical deviation strictly as a between-group relative comparison under identical scenario conditions, rather than as an absolute measure of vertical-path-tracking accuracy or procedural compliance. I used the along-track distances from the ideal path to the waypoints and linearly interpolated altitude between consecutive waypoint constraints. Let *s* represent along-track distance. Then ideal vertical reference profile can be written as shown in Equation 24

$$h_{ideal}(s) = linear\ interpolation\ of\ waypoint\ altitude\ constraints \qquad \text{Equation 24}$$

where $h_{ideal}(s)$ gives the ideal altitude in feet MSL at along-track distance *s*.

For each aircraft position, I compared the aircraft latitude and longitude with the ideal lateral path. The pipeline computed the great-circle distance between the aircraft position and each point on the ideal path using the haversine distance (Agramanisti Azdy & Darnis, 2020). The nearest ideal path point defined the lateral reference point for that aircraft sample. For aircraft sample i, the aircraft position was (lat_i, lon_i) and each ideal path point was $\left(lat_j^{ideal}, lon_j^{ideal}\right)$. The haversine distance between the aircraft and j^{th} ideal path point is given by Equation 25

$$d_{ij} = 2R\ arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_i)\cos\left(\phi_j\right)\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right) \qquad \text{Equation 25}$$

where R is the Earth's radius, ϕ is latitude in radians, λ is longitude in radians, $\Delta\phi$ is the latitude difference, and $\Delta\lambda$ is the longitude difference.

The nearest ideal path point is given by Equation 26

$$j^* = arg \min_j d_{ij} \quad \text{Equation 26}$$

I then computed cross-track error as shown in Equation 27

$$CTE_i = d_{ij^*} \quad \text{Equation 27}$$

where CTE_i is the lateral deviation magnitude in meters.

This study used unsigned cross-track error because the primary research question focused on the magnitude of deviation rather than its left/right direction. Both leftward and rightward deviation represent reduced tracking accuracy relative to the desired RNAV centerline; therefore, unsigned CTE was appropriate for evaluating path-following accuracy.

After identifying the nearest ideal path point, I assigned each aircraft sample an along-track distance. If the nearest ideal point was *j**, then the aircraft sample's along-track position is given by Equation 28

$$s_i = s_{j^*}^{ideal} \quad \text{Equation 28}$$

where s_i is the aircraft sample's assigned along-track distance, and $s_{j^*}^{ideal}$, is the cumulative distance of the nearest ideal path point. This along-track assignment served three purposes. First, it enabled the pipeline to identify the approach segment corresponding to each aircraft sample. Secondly, it estimated the ideal altitude at the same along-track position. Third, it allowed the distance-normalized analysis to group samples into fixed along-track distance bins.

I computed vertical deviation by comparing the actual aircraft altitude with the ideal altitude at the aircraft assigned along-track position. For the aircraft sample *i*, the actual altitude was h_i and the ideal altitude at the assigned along-track position was $h_{ideal}(s_i)$. The vertical deviation is then given by Equation 29

$$VDev_i = h_i - h_{ideal}(s_i) \qquad \text{Equation 29}$$

where $VDev_i$ is in feet, whose sign convention is given below:

$$V\ Dev_i > 0\ \gg aircraft\ above\ the\ ideal\ path\ profile$$

$$V\ Dev_i < 0\ \gg aircraft\ below\ the\ ideal\ path\ profile$$

Because the aircraft were initialized below the intended early-approach altitude, many vertical deviation values were negative. So, signed vertical deviation should be interpreted as a relative comparison between groups, not as a direct or absolute measure of pilot procedural error.

I later combined three-dimensional deviation metrics to summarize lateral and vertical path errors into a single value. Because lateral deviation was measured in meters and vertical deviation was measured in feet, vertical deviation was therefore converted to meters as shown in Equation 30.

$$V\ Dev_{i,m} = 0.3048\ \times V\ Dev_i \qquad \text{Equation 30}$$

Then, three-dimensional deviation can be written as shown in Equation 31

$$Dev\ 3D_i = \sqrt{CTE_i^2 + V\ Dev_{i,m}{}^2} \qquad \text{Equation 31}$$

where $Dev3D_i$ is the combined path deviation in meters. Lower $Dev3D$ values indicate that the aircraft remained closer to the ideal three-dimensional reference path.

The first stage of the analysis computed deviation metrics at every X-Plane sample. These sample-level values were then averaged over time to produce mean, RMS, maximum, and standard deviation metrics. For example, the time weighted RMS cross-track error is given by Equation 32.

$$CTE_{RMS,time} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} CTE_i^2} \qquad \text{Equation 32}$$

where N is the number of simulator samples.

This metric describes the deviation experienced across the recorded task period. However, it is time-weighted. If one pilot flew more slowly, paused, or spent more time in a particular region of the approach, that region contributed more strongly to the final metrics. Therefore, time-weighted metrics can be influenced by flight duration, groundspeed, sampling rate, or time spent in a specific portion of the approach. Since my aim is to evaluate path-following accuracy along the approach path, I added distance-normalized metrics and used them as the primary performance outcomes.

Distance-normalized metrics gave equal weights to equal distances along the approach path. First, I divided the ideal approach path into uniform along-track distance bins. In the modified analysis, I used a 50 m grid spacing. This created a sequence of along-track bins $[s_0, s_1], [s_1, s_2], \dots., [s_{K-1}, s_K]$ where each bin covered approximately 50 m of the ideal path.

Second, I assigned each aircraft sample to a bin based on its along-track position s_i. For each bin k, I collected all aircraft samples assigned to that distance interval.

Then, I computed bin-level deviation metrics. For example, the RMS cross-track error for bin k is given by Equation 33.

$$CTE_{RMS,k} = \sqrt{\frac{1}{n_k}\sum_{i \epsilon k} CTE_i^2} \qquad \text{Equation 33}$$

where n_k is the number of aircraft samples in bin *k.*

Finally, the analysis computed participant-level distance-normalized metrics across valid bins. For RMS lateral deviation, the metric is given by Equation 34

$$TE_{DN,RMS} = \sqrt{\frac{1}{K_v}\sum_{k=1}^{K_v} CTE_{RMS,k}^2} \qquad \text{Equation 34}$$

where K_v is the number of valid distance bins.

Vertical deviation and three-dimensional deviation used the same approach as in Equation 34. This method reduced bias from differences in how long each pilot spent in different regions of the approach. Therefore, I treated the distance-normalized metrics as the primary flight-path performance outcomes.

The Minimum Decision Altitude (MDA) used in the analysis was MDA: 8380 ft MSL. For each participant, the percentage below the MDA is calculated by

$$Below\ MDA\ (\%) = \frac{N(h_i < MDA)}{N} \times 100 \qquad \text{Equation 35}$$

where $N(h_i < MDA)$ is the number of samples with aircraft altitude below the MDA, and N is the total number of valid altitude samples. The simulation required the participants to make a go-around decision. Therefore, the analysis did not treat below-MDA exposure as a standalone procedure violation or as a simple failure metric. Instead, I used it as a descriptive safety-related indicator that helped quantify how much of the recorded trajectory occurred below the published MDA.

6.6.2 Statistical Analysis

Each participant contributes one value per metric. I did not treat individual flight samples or distance bins as independent observations because those observations were repeated measurements from the same participants. Similar to the previous analysis (HRV + Pupil diameter), the Mann-Whitney U test is used. I reported the mean, standard deviation, median, Interquartile Range (IQR), test statistics, p-value, and rank-biserial effect size. The rank biserial effect size was coded so that positive values indicate higher values in the tailored interface group and negative values indicate higher values in the baseline group. For deviation metrics, lower values generally indicate better performance; therefore, the direction of the effect size must be interpreted in light of the metric's definition. The definition and interpretation of each metric are provided in Table 19.

Table 19 Flight-path-deviation metric definitions used for the analysis

Metric	**Definition**	**Interpretation**
CTE DN RMS	Distance-Normalized RMS cross-track error	Lower values indicate better lateral path tracking
CTE DN Mean	Distance-normalized mean cross-track error	Lower values indicate smaller average lateral deviation
VDev DN RMS	Distance-normalized RMS vertical deviation	Lower values indicate better vertical path tracking
VDev DN Mean	Distance-normalized signed mean vertical deviation	Negative values indicate flight below the revised ideal profile

3D DN RMS	Distance-normalized RMS combined lateral and vertical deviation	Lower values indicate better total path tracking
Below MDA %	Percentage of valid samples below MDA	Lower values indicate less time below MDA
Path Coverage %	Percentage of distance bins with valid aircraft samples	Used to check the comparability of the analyzed path length

Since the sample size of the tailored interface is small and the experiment is a between-subject design, I interpreted the results as directional performance evidence rather than conclusive statistical evidence.

Table 20 Distance Normalized flight path deviation results. Lower values indicate better performance. Vertical deviation mean is signed; negative values indicate flight below the revised ideal vertical reference profile

Metric	**Baseline Mean ±SD**	**Baseline Median [IQR]**	**Tailored Mean ±SD**	**Tailored Median [IQR]**	**U**	**P**	**r**
CTE DN, RMS	851.6 ± 829.6	531.3 [211.6]	536.8 ± 61.2	544.29 [41.4]	107	0.74	-0.08
CTE DN Mean, m	490.2 ± 706.6	256.6 [149.2]	246.6 ± 36.7	245.1 [19.9]	115	0.50	-0.16
VDev DN RMS, m	361.7 ± 65.8	348.7 [82.0]	331.2 ± 69.6	336.1[21.8]	122	0.32	-0.23
VDev DN Mean, m	-274.2 ± 67.7	-253.1 [93.5]	-237.4 ± 68.2	-230.4 [51.5]	67	0.17	0.32
3D DN RMS, m	949.8 ± 802.7	647.4 [236.5]	633.0 ± 73.0	655.9 [84.4]	120	0.37	-0.21
Below MDA, %	19.9 ± 11.0	19.9 [4.8]	19.4 ± 5.8	21.0 [3.6]	87	0.61	0.12
Path Coverage	53.7 ± 8.7	57.5 [4.4]	56.5 ± 2.5	57.6 [0.5]	97	0.94	0.02

As shown in Table 20, the tailored-interface group showed a lower average distance-normalized lateral deviation than the baseline group. The baseline group had a CTE DN RMS of 851.6±829.6 m, whereas the tailored interface group had a CTE DN

RMS of 536.8±61.2 m. The median values were closer, with the baseline group at 531.3 m and the tailored-interface group at 544.3 m. This difference between the mean and median indicates that the baseline was strongly influenced by extreme cases of lateral deviation. The group comparison did not reach statistical significance ($U = 107$, $p = 0.74$, $r = -0.08$).

The distance-normalized mean cross-track error showed a similar directional pattern. The baseline group had a CTE DN Mean of 490.2±706.6 m, with a median of 256.6 m. The tailored-interface group had a CTE DN Mean of 246.6±36.7 m and a median of 245.1 m. The tailored group showed a much narrower distribution, suggesting more consistent lateral path tracking. However, the group comparison did not reach statistical significance ($U = 115$, $p = 0.50$, $r = -0.16$).

Vertical deviation favored the direction of the tailored-interface group. The baseline group had a VDev DN RMS of 361.7±65.8 m, while the tailored interface group had a VDev DN RMS of 331.2 ± 69.6 m. This comparison did not reach statistical significance ($U = 122$, $p = 0.32$, $r = -0.23$). The signed mean vertical deviation showed that both groups flew below the ideal vertical profile. The baseline group had a mean vertical deviation of -274.2 ± 67.7 m, while the tailored-interface group had a mean vertical deviation of -237.4 ± 68.2 m. This indicates that the tailored interface group was less far below the revised reference profile on average, although the comparison did not reach statistical significance ($U = 67$, $p = 0.17$, $r = 0.32$). Because the scenario initialized participants at a lower-than-intended rate, this result should be interpreted as a relative group comparison rather than a direct measure of procedural compliance.

The combined three-dimensional deviation also favored the tailored-interface group. The baseline group had a 3D DN RMS of 949.8±802.7 m, whereas the tailored interface group had a 3D DN RMS of 633.0±73.0 m. The median values were 647.4 m for baseline and 655.9 m for tailored. Similar to the CTE DN RMS result, the large baseline mean and standard deviation indicate that extreme baseline cases increased the group average. The tailored-interface group showed a narrower distribution, suggesting more consistent three-dimensional path tracking. However, the group comparison was not statistically significant ($U = 120$, $p = 0.37$, $r = -0.21$).

The percentage of samples below MDA was similar between groups. The baseline group spent 19.9±11.0% of samples below MDA, while the tailored-interface group spent 19.4±5.8% below MDA. The median values were also similar, 19.9% for the baseline and 21.0% for tailored interface group. This result suggests that the tailored interface did not produce a clear difference in the MDA exposure below.

Path coverage was also similar between the two groups. The baseline group covered 53.7±8.7% of distance-normalized bins, while the tailored interface groups covered 56.5±2.5%. The medians were 57.5% and 57.6%, respectively. The group comparison was not statistically significant ($U = 97$, $p = 0.94$, $r = 0.02$). This similarity supports the comparability of the distance-normalized result, although both groups covered only about half of the full REDX-to-RW27 reference path.

In conclusion, the flight-path-deviation results suggest that the tailored interface group exhibited lower average lateral, vertical and 3D deviations and narrower distributions in several metrics. However, none of the comparisons reached $p < 0.05$, so the results should be interpreted as directional evidence rather than conclusive statistical evidence.

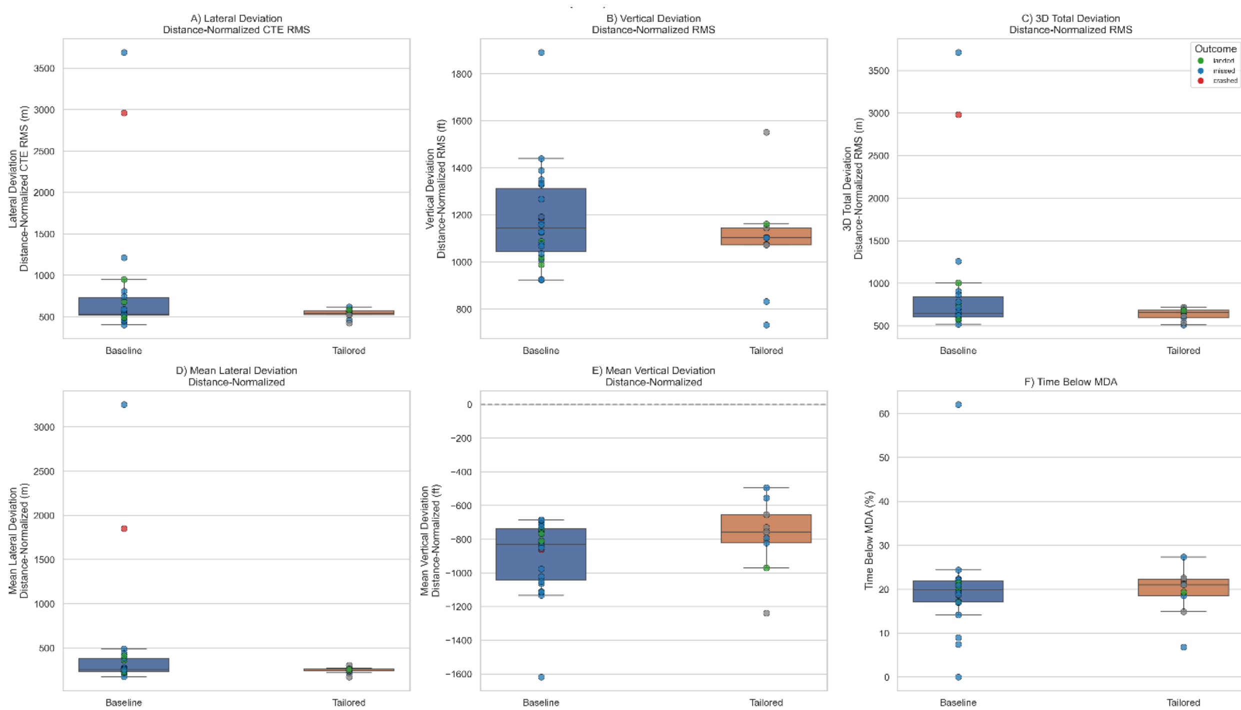

Figure 28 Distance-Normalized flight-path-deviation metrics for baseline and tailored-interface group.

Figure 28 shows that the baseline group exhibited greater variability, particularly in lateral and 3D deviations. This supports the interpretation that the tailored interface may have produced more consistent path-following behavior, even though the statistical comparisons did not reach significance.

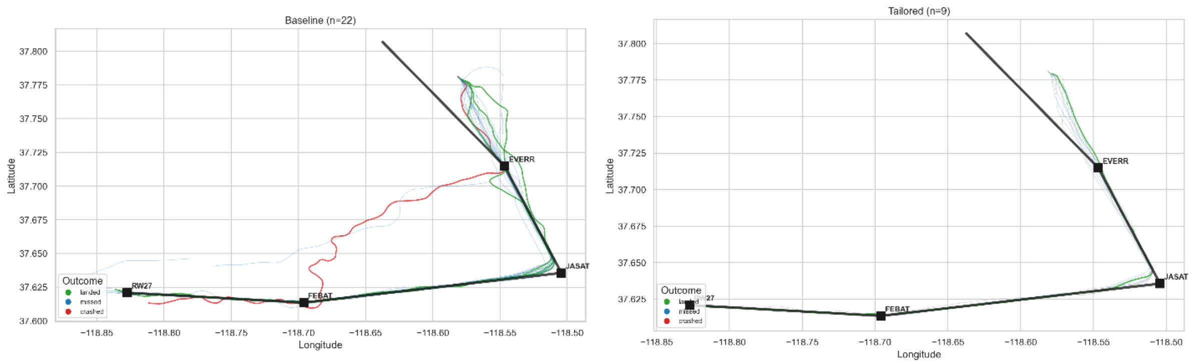

Figure 29 Participants flight trajectories compared with the ideal lateral path. The black line represents the ideal RNAV lateral path, and the colored traces represent participant trajectories. Outcome colors indicate whether the participant landed, executed a go-around, or crashed. Here, the go-around decision is interpreted as part of the intended operational context rather than a failed trial

Figure 29 illustrates the high variability in baseline lateral deviation metrics. Several baseline trajectories deviated substantially from the ideal approach path, while the tailored-interface trajectories were more tightly clustered. This figure is important because it shows distributional behavior that is not fully captured by p-values.

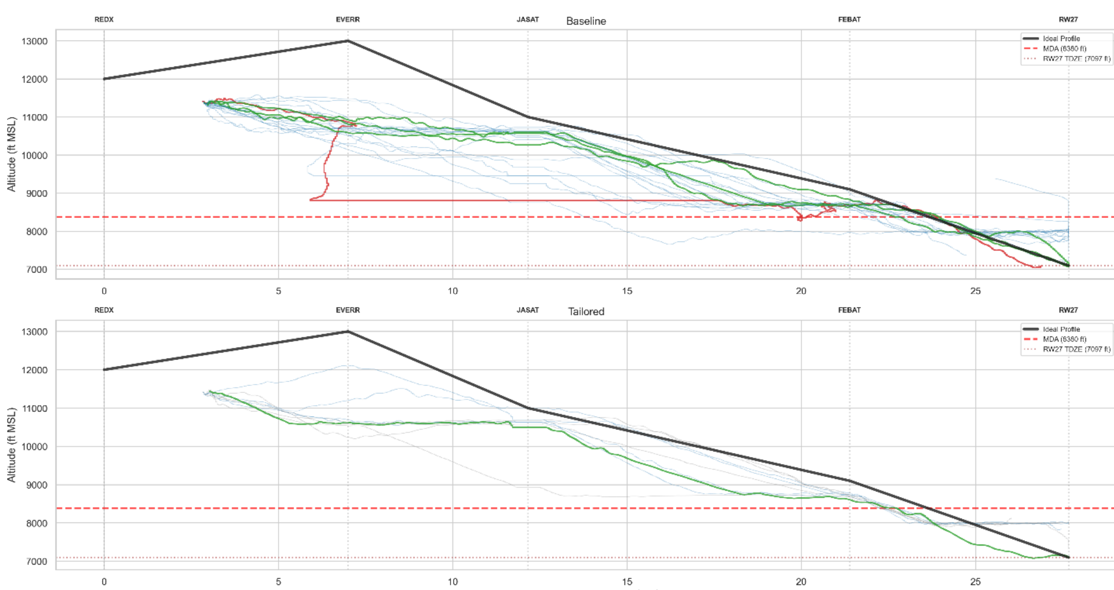

Figure 30 Participants altitude profiles compared with the ideal vertical reference profile. The black line represents the revised ideal vertical profile based on the intended approach altitude constraints. The dashed red line represents the MDA, and the dotted line represents the runway elevation

Figure 30 shows that both groups were below the ideal vertical profile during much of the approach. This pattern reflects both flight-path behavior and the scenario-design limitation. Therefore, the vertical profile results should be interpreted as relative group differences under the same simulation setup rather than as absolute procedural compliance.

6.6.3 Discussion

The flight-path deviation analysis provides evidence of directional operational performance. The tailored interface group showed lower average distance-normalized lateral, vertical and 3D deviation than the baseline group. The tailored-interface group also showed a narrower distribution for lateral and 3D deviation, whereas the baseline group contained several extreme deviation cases. This pattern suggests that the tailored interface may have supported more consistent path-following behavior.

This vertical deviation result is obtained from the published vertical profile. However, the experiment scenario initialized participants at a lower altitude than intended for the early approach. As a result, both groups showed negative mean vertical deviation relative to the reference profile. None of the flight path deviation comparisons reached the conventional $p < 0.05$ significance threshold. However, the absence of statistical significance should not be interpreted as evidence that the interface and the baseline (without interface) were equivalent. The tailored group remained smaller than the baseline group, which limited statistical sensitivity. Therefore, this study emphasizes the direction of effects, effect sizes, robust summaries, and trajectory visualization rather than relying only on p-values.

The results align with the broader pattern observed in previous sections. The subjective analysis suggested that the tailored interface improved perceived performance and situation understanding. The HRV analysis suggested greater physiological engagement in the tailored-interface group. The pupil dilation analysis did not show a clear difference in pupil-based visual cognitive effort. The flight path deviation analysis suggests that the tailored interface group may have flown more consistently and closer to the desired path, especially in vertical and combined three-dimensional deviation.

Overall, the flight path deviation analysis suggests that the tailored interface may have provided operational support during the approach. Nonetheless, the evidence remains preliminary due to the small, tailored interface size and high baseline variability. A larger, more balanced sample would be needed to determine whether the observed directional trends reflect a reliable interface effect.

This analysis shows whether the pilots maintained the desired approach path, but it does not explain how pilots used visual information while performing the task. A pilot may fly closer to the desired path because the interface improved information availability, because the pilot scanned instruments more effectively, or because the pilot relied more heavily on specific visual cues. Therefore, the next section examines the visual attentional behavior using gaze path analysis and heat maps. These analyses evaluate where pilots gazed during the task and whether attention allocation differed between the baseline and tailored-interface groups. The visual-attention results help explain subjective, physiological, and flight performance patterns by showing how pilots distributed their attention across the outside scene, flight instruments, and tailored interface information.

6.7 Visual Attention

The visual attention analysis examined how pilots distributed gaze across major cockpit information sources during the approach. While flight-path-deviation analysis evaluated aircraft path-following performance, the visual attention analysis evaluated where pilots allocated visual attention while monitoring aircraft state, navigation information, outside visual cues, and the tailored interface. The main purpose of this section was to determine whether the tailored interface changed pilots' allocation of visual attention across cockpit information sources.

6.7.1 Areas of Interest

I processed the eye-tracking data using a custom Python-base gaze-analysis pipeline developed in-house by the ÆON lab. The pipeline used the XDF recording, a cockpit reference image, and a predefined Areas of Interest (AOI) configuration file as inputs. During the analysis, AOI is defined by a rectangular region on the cockpit reference image using normalized image coordinates. This format allowed the same AOI boundaries to remain tied to the reference image during analysis. The processing pipeline loaded and time-synchronized the gaze and head-pose streams, projected gaze directions onto the cockpit reference image. The pipeline assigned each valid gaze sample to an AOI, detected fixations using an I-VT velocity threshold method, and exported AOI-level visual-attention metrics. The exported metrics included dwell percentage, fixation count, mean fixation duration, revisits, time to first fixation and transition related measures.

Both groups shared the following AOI: the G1000 altitude indicator, altitude tape, rate-of-climb tape, multi-function display, course deviation indicator, and outside visual scene. The tailored-interface group had one additional AOI representing the entire experimental interface.

Table 21 Visual attention AOI definitions

	AOI Label	Description	Interpretation
1	ATI	G1000 attitude indicator	Aircraft attitude/state monitoring
2	ALT	Altitude tape	Vertical-state monitoring
3	ROC	Rate-of-climb tape	Climb/descent performance monitoring
4	CDI	Course deviation indicator	Lateral course/path tracking
5	SPD	Speed tape	Airspeed Monitoring
6	MFD	Multi-function display	Map/navigation monitoring
7	Outside	Outside visual scene (Front)	External visual-reference monitoring
8	Interface	Whole tailored decision-support interface	Attention to the experimental interface as a complete display

The tailored interface combined several decision-relevant elements into one display region, including performance and alert, AFM climb performance requirements, terrain-relative altitude information, aircraft state values, upcoming waypoint constraints, map/path information, and lateral cross-track error. Therefore, the AOI interface represented the entire tailored display region rather than a single, isolated instrument. Any dwell time on the interface AOI indicates attention to the integrated decision-support display as a whole.

Figure 31 Different instruments including the tailored interface constituted various Areas of Interest.

6.7.2 Metric Selection

The baseline eye-tracking system sampled data at approximately 4 Hz, whereas the tailored-interface system sampled data at approximately 100 Hz. This difference in sampling frequency strongly affects fixation-based measures. Higher-frequency gaze data can detect more short-duration fixations and more frequent transitions, whereas lower-frequency data can merge visual samples into longer apparent fixations. For this reason, this analysis did not use raw fixation count, total transition count, or mean fixation duration as primary group-comparison outcomes. Instead, it used AOI dwell percentage as the primary visual attention metric. The dwell percentage is defined as given in Equation 36

$$Dwell\ Percentage\ (AOI) = \frac{Time\ spent\ in\ that\ AOI}{Total\ \ gaze\ time} \times 100 \qquad \text{Equation 36}$$

The analysis selected dwell percentage because it represents the proportion of visual attention allocated to each AOI and is less sensitive to sampling-frequency differences than raw fixation count. The analysis also converted fixation count into participant-level fixation share as shown in Equation 37

$$Fixation\ share\ (AOI) = \frac{Fixation\ count\ in\ that\ AOI}{Total\ fixation\ count\ across\ AOIs} \times 100 \qquad \text{Equation 37}$$

Because the pipeline still detected fixation events from data streams with different temporal resolutions, this section treats fixation-share results as secondary supporting evidence.

6.7.3 Results

Table 22 presents the AOI dwell-percentage results for the common cockpit AOIs. The tailored group spent a greater percentage of dwell time on the attitude indicator, altitude tape, multi-function display, and outside scene, and a smaller percentage on the rate-of-climb tape, speed tape, and course deviation indicator.

Table 22 Dwell percentage of baseline and tailored group participants in the defined AOI

AOI	**Baseline Mean ± SD**	**Baseline Median [IQR]**	**Tailored Mean ± SD**	**Tailored Median [IQR]**
ATI	20.53 ± 11.23	18.30 [13.30]	22.60 ± 11.04	21.00 [17.15]
ALT	5.06 ± 3.06	4.15 [4.38]	9.70 ± 7.48	7.20 [11.85]
ROC	3.45 ± 3.65	2.50 [3.73]	1.54 ± 1.20	1.70 [1.40]
SPD	2.11 ± 2.18	1.30 [2.38]	0.99 ± 1.00	1.00 [1.13]
MFD	4.32 ± 3.75	3.20 [2.53]	4.53 ± 3.31	3.60 [3.25]

CDI	13.50 ± 7.74	14.05 [10.90]	9.34 ± 6.52	8.65 [10.20]
Outside	3.10 ± 3.68	1.50 [4.05]	4.86 ± 2.85	5.15 [3.23]

The most prominent pattern was the increase in altitude tape dwell percentage in the tailored group. Pilots in the tailored group spent 9.70% of their dwell time on altitude tape, compared with 5.06% in the baseline group. This finding suggests that pilots using the tailored interface devoted more visual attention to vertical state monitoring. Pilots in the tailored group also spent slightly more dwell time viewing outside. Because the scenario was an IFR approach, the outside scene dwell may reflect brief checks for external visual cues, attempts to acquire the runway environment near the decision portion of the approach, or gaze behavior related to the simulator visual scene and AOI masking. Pilots in the tailored group spent less dwell time on the CDI, ROC tape, and speed tape. This pattern should not be interpreted as a simple neglect of lateral or performance limitation. The tailored interface itself displayed lateral cross-track error, current speed, and rate of climb, altitude, recommended climb speed and rate, and upcoming waypoint altitude. Therefore, lower dwell on individual G1000 elements may indicate that some lateral-path and performance information was accessed through the integrated interface rather than through separate conventional instrument AOIs.

The interface AOI represented the entire tailored decision-support display, not a specific subcomponent. The interface displayed alert status, terrain-relative altitude information, AFM climb performance requirements, current speed, rate of climb, altitude, upcoming waypoint constraints, and lateral cross-track error (XTK). Therefore, dwelling on this AOI indicates attention to the integrated decision-support display as a whole, but it does not reveal which specific interface elements the pilot used.

Table 23 Tailored-interface AOI descriptive metrics

Metric	**Mean ± SD**	**Median [IQR]**
Interface dwell percentage	4.74 ± 3.59	4.00 [2.75]
Interface fixation count	50.33 ± 40.03	37.00 [34.25]
Interface mean fixation duration, ms	526.69 ± 89.18	514.45 [114.10]

Pilots in the tailored group spent a modest proportion of their dwell time on the interface. This result is consistent with the intended role of the interface as a supplementary decision-support aid rather than a primary flight display. The interface was designed to support information processing by consolidating critical cues that would otherwise need to be gathered from multiple sources and integrated in working memory. Therefore, the result suggests that the pilot incorporated the interface into their visual scan while continuing to rely primarily on conventional flight instruments for aircraft control.

Since the baseline eye-tracking data and tailored group data were sampled at different frequencies, which can strongly affect raw fixation counts, the data were converted to percentages of each participants fixations allocated to each common AOI using Equation 37. Table 24 presents the fixation-share results for the common AOIs.

The fixation-share results generally support the dwell-percentage findings. Pilots in the tailored group allocated a larger share of fixations to the attitude indicator and altitude tape. The altitude-tape differences were especially notable: the tailored group allocated 19.32% of common AOI fixation to ALT, compared with 7.33% in the baseline group. This pattern suggests greater visual sampling of vertical state information in the tailored group.

Table 24 Fixation share results for common AOIs

AOI	Baseline Mean ± SD	Baseline Median [IQR]	Tailored Mean ± SD	Tailored Median [IQR]
ATI	39.73 ± 19.32	36.76 [18.14]	42.98 ± 13.62	42.09 [18.46]
ALT	7.33 ± 5.17	7.92 [9.21]	19.32 ± 19.93	12.51 [16.58]
ROC	3.76 ± 4.75	2.21 [4.81]	1.61 ± 1.18	1.74 [1.66]
SPD	2.37 ± 3.67	0.86 [2.52]	1.90 ± 2.40	1.14 [1.90]
MFD	13.00 ± 9.22	10.26 [10.08]	7.58 ± 6.19	6.09 [4.13]
CDI	25.19 ± 15.24	24.01 [21.80]	17.49 ± 12.99	17.39 [16.74]
Outside	8.62 ± 14.50	3.69 [8.66]	9.13 ± 6.63	8.23 [9.87]

The tailored group allocated a smaller share of the fixation to the CDI and MFD. The tailored interface itself displayed lateral cross-track error, route/path information, and upcoming waypoint information. Therefore, the lower fixation share on the CDI and MFD may indicate that some lateral path and navigation information was accessed through the integrated decision-support display rather than solely through the conventional G1000 displays. Similarly, lower fixation was observed on the ROC and speed tape. The tailored interface displayed current speed, rate of climb, and recommended climb speed and rate. Therefore, pilots may have obtained some performance-related information from the interface rather than from separate conventional tape elements.

I used scanpath entropy to describe the dispersion and complexity of visual scanning across AOIs. Stationary entropy describes how broadly gaze was distributed across AOIs, while transition entropy describes the variability of movement between AOIs. Higher entropy generally indicates a more distributed or varied scan pattern, whereas lower entropy indicates a more concentrated or repetitive scan pattern. The groups showed similar normalized stationary entropy values. The baseline group had a normalized stationary entropy of 0.685±0.163, while the tailored group had 0.708±0.096. The groups also showed similar normalized transition entropy values, with the baseline group at 0.645±0.154 and the tailored group at 0.673±0.089. These results suggest that the tailored interface did not substantially increase or decrease overall scan dispersion.

In addition, the tailored group visited more AOIs on average, with 7.50±0.67 compared with 6.18±0.85 in the baseline group. This difference was partly expected because the tailored group had access to an additional AOI, i.e., the interface. This indicates that the pilots in the tailored group sampled a broader set of available information. Overall, the entropy and AOI-coverage results suggest that the tailored interface was integrated into the scan pattern without fundamentally reorganizing gaze behavior.

6.7.4 Discussion

The visual attention results suggest that pilots incorporated the tailored interface into their scan patterns in a manner consistent with its intended purpose. The interface was designed to support decision-making by improving information processing in a complex mountainous environment. Specifically, it was intended to reduce the need for pilots to perform complex task analysis in working memory, consolidating critical information related to terrain, aircraft performance, path deviation, and altitude constraints. To verify that the sampling rate difference did not drive the visual attention results, the tailored interface gaze data were down sampled to 4 Hz to match the baseline recordings, and all dwell-percentage outcomes were recomputed. Although the numerical values shifted slightly, the direction and pattern were unchanged, confirming the dwell percentage is robust to sampling frequency difference.

Pilots using the tailored interface spent a modest proportion of dwell time on it while continuing to devote most of their attention to conventional flight instruments, which supports the pilot’s decision process while preserving the normal instrument scan required for IFR flight. The strongest visual-attention finding was the shift toward vertical state monitoring. Pilots in the tailored group showed higher dwell percentage and higher normalized fixation share on the altitude tape. This pattern aligns with the operational demands of the IFR go-around decision context, where pilots needed to determine whether altitude, aircraft performance, terrain clearance, and lateral path position supported continuing the approach. The tailored group also showed lower dwell time on the CDI, ROC, and SPD tape. The tailored interface displayed several related values, including lateral cross-track error, speed, rate of climb, altitude, recommended climb speed and rate, and the altitude of the upcoming waypoint. Therefore, the reduction in dwell time on individual G1000 elements may reflect partial redistribution of attention toward the integrated interface display. In this sense, the interface may have reduced the need for pilots to gather and mentally combine the same information from separate instruments.

Since the experiment involved an IFR approach, the slightly higher outside-scene dwell in the tailored group suggests brief attempts to acquire external visual references near the approach's decision phase. The entropy results suggest that the tailored interface did not fundamentally reorganize visual scanning behavior. Instead, pilots appeared to have integrated the interface into the existing instrument scan as an additional source of consolidated decision-support information, consistent with the interface design concept.

Taken together, the visual attention findings suggest that the tailored interface supported attentional allocation by adding an integrated decision-support source rather than replacing conventional instrument scanning. In conclusion, the tailored interface was used modestly for its intended purpose, with the clearest shift in attention toward vertical-state monitoring.

6.8 Overall Discussion

The combined results suggested that the tailored interface supported pilot information processing during the mountainous approach. The subjective results showed the clearest directional benefit. The pilot who used the tailored interface reported lower raw and weighted NASA-TLX workload and frustration, better perceived performance, and higher SART understanding. These results support the central purpose of the interface, to help pilots interpret task-critical information without forcing them to mentally integrate terrain, altitude, path deviation, and performance cues from separate sources.

HRV showed a larger rest-to-task reduction in lnRMSSD in the tailored-interface group, suggesting stronger task-related physiological engagement. However, this pattern did not align with higher frustration or poorer perceived performance. Instead, pilots reported better understanding and lower workload. Therefore, the HRV results likely reflect active engagement with the tasks and interface, rather than a simple increase in stress. Mean heart rate did not distinguish the groups, further supporting the value of HRV as a more sensitive physiological measure.

Pupil dilation provided neutral evidence. The tailored interface did not produce a clear difference in raw pupil diameter, pupil variability, pseudo-baseline-z-score, or percent change from the pseudo-baseline. This result suggests that the interface did not substantially increase pupil-based visual-cognitive effort during the period of the analyzed task. The null pupil result does not contradict the subjective or HRV findings because the pupil diameter responds to luminance, gaze direction, headset fit, arousal, and individual physiology in addition to workload. Instead, the pupil result suggests that the tailored interface may have increased task engagement without producing a measurable increase in pupil dilation.

The flight-performance results also favored the tailored interface in a directional manner. The tailored-interface group showed lower average vertical and combined 3D deviation and more consistent path tracking, although the comparisons did not reach statistical significance. The trajectory plots and robust summaries suggest that the baseline group contained more extreme deviations, while the tailored-interface group showed a tighter performance distribution. Therefore, the interface may have helped pilots maintain a more stable approach path under demanding terrain and performance conditions.

The visual-attention results explain how the interface may have supported the performance. Pilots in the tailored-interface group paid measurable attention to the interface itself and showed increased attention to altitude-related information. At the same time, some conventional display areas received a normalized share of fixation. This pattern suggests that the interface changed the distribution of visual attention rather than simply adding another display to monitor. The interface may have helped pilots access critical altitude, terrain, and path information more directly, reducing the need to search across multiple cockpit sources.

The study did not include a formal qualitative protocol. However, during debriefing several tailored interface participants commented on the interface when asked about the approach. Two participants noted that the upcoming waypoint information reduced the need to refer to the approach chart during the flight, and one participant highlighted the altitude-specific performance information. These informal observations were not systematically collected or coded and are reported only as supplementary context. However, they are consistent with the direction of the quantitative subjective results.

Taken together, the findings suggest that the tailored interface acted as a decision-support aid. It appeared to support comprehension, reduce frustration, and improve consistency while keeping the pilot actively engaged. This interpretation aligns with this study framework, i.e., to reduce unnecessary working-memory integration by presenting mission-critical information in a more salient and context-relevant form. Nevertheless, this study interprets the findings as preliminary. Several comparisons showed meaningful directional trends and moderate effect sizes, but most did not reach statistical significance. The small, tailored-interface sample, between-subjects design, lack of a true resting-pupil baseline, and scenario-design limitations constrain the strength of the conclusions.

Overall, Chapter 6 shows that the tailored interface produced a consistent pattern of directional benefits across subjective workload, situation awareness, operational performance, and visual-attention behavior, while physiological measures suggested active task engagement rather than passive workload reduction. The next chapter summarizes these findings, states the main contributions of the dissertation, discusses study limitations, and provides recommendations for future interface development, experimental validation, and operational implementation.

7 Conclusion

This dissertation developed and evaluated a context-sensitive cockpit decision-support framework for complex flight operations. This framework can apply to different operational contexts by adjusting which information elements receive priority while preserving the same underlying design logic. This study used mountainous CFIT-risk operations for validation. The framework prioritized terrain, altitude, aircraft performance, and flight-path deviation because these elements directly support pilot decision-making during high-workload approach operations.

The work addressed a central problem in cockpit information processing: pilots often have access to the required information, but the information is not always organized according to operational risk or cognitive demand. During mountainous approach operations, pilots must integrate terrain clearance, altitude constraints, aircraft performance limits, flight-path deviation, weather, and navigation information. This integration increases working memory demand and can weaken situational awareness. The proposed framework addresses this issue by identifying context-critical information and presenting it with greater salience, enabling pilots to interpret the situation more efficiently.

The study combined accident-analysis evidence, human factors theory, interface design, and flight-simulation evaluation to reduce cognitive overload and improve information processing. The generic interface design portion used the concept of human factors and display design guidelines to create a fixed zone architecture. The accident-analysis portion identified CFIT-related risk factors and information-processing demands and translated those findings into a tailored cockpit display. The evaluation portion tested whether the tailored interface influenced subjective workload, situation awareness, physiological response, pupil dilation, flight performance, and visual attention.

The results provide preliminary support for the proposed framework. While the tailored interface did not reduce the workload of the task, it appeared to help pilots manage workload more effectively. The NASA-TLX results showed lower perceived workload, frustration, and performance-related workload, and higher perceived performance success in the tailored-interface group. The SART results showed higher overall situation awareness and stronger situational understanding. These findings support the interface's main purpose: helping pilots interpret task-critical information rather than forcing them to search for, remember, and integrate it manually.

The physiological results provided a more detailed interpretation. HRV showed a greater rest-to-task reduction in lnRMSSD for the tailored-interface group, suggesting stronger task-related physiological engagement. However, the tailored-interface group also reported lower frustration and higher situational understanding. Therefore, the HRV result likely reflects active engagement with the task and interface, rather than a simple increase in stress. Mean heart rate did not distinguish the groups, suggesting that HRV captured a more sensitive workload-related response than heart rate alone.

The pupil dilation results did not show a clear difference between groups. Raw pupil diameter, pupil variability, pseudo-baseline-normalized pupil change, and percent change from pseudo-baseline remained similar across the baseline and tailored-interface groups. This result suggests that the tailored interface did not produce a measurable increase in pupil-based cognitive effort during the period of the analyzed task. The neutral pupil result does not contradict the subjective or HRV findings because pupil diameter responds to cognitive and non-cognitive factors, including display luminance, gaze direction, headset fit, arousal, and individual physiology.

The flight performance results also favored the tailored interface in a directional manner. The tailored interface group showed lower average vertical deviation, lower combined three-dimensional deviation, and more consistent path-tracking behavior. The baseline group showed greater variability and included more extreme deviation cases. Although these results did not reach statistical significance, they suggest that the tailored interface may have supported more stable operational performance during the mountainous approach.

The visual attention results helped explain how pilots used the interface. Pilots in the tailored interface group allocated measurable attention to the interface and showed increased attention toward altitude-related information. This pattern suggests that the interface changed how pilots accessed and used critical information during the approach. Rather than simply adding another display, the interface appeared to support a different scan strategy that directed attention toward operationally important cues.

Overall, the findings support the central argument of this dissertation: flight deck safety depends not only on information availability but also on how the interface organizes, prioritizes, and presents that information under operational demands. The proposed framework offers a method for aligning information salience with context-specific risk. In the mountainous CFIT validation context, the tailored interface produced directionally favorable outcomes across subjective workload, situation awareness, flight performance, and visual attention while maintaining active pilot engagement.

These findings remain preliminary because of the small tailored-interface sample, the between-subjects design, and several data-quality constraints. However, the convergence across multiple measures suggests that the framework offers a promising direction for cockpit decision support design. The dissertation contributes more than a single display prototype. It provides a generalizable approach to identifying critical information in a given operational context and presenting it in a way that supports pilot cognition, situation awareness, and decision-making.

7.1 Limitations

A key limitation concerns the formative validation of the tailored-interface design. I developed the interface using human factors principles, aviation display design guidelines, and information priorities identified through the complex network theory. However, I did not ask subject matter experts to review the design or use representative pilots to evaluate it. I collected eye-tracking data during the experimental evaluation but did not use the interface to optimize or re-design the tailored interface. Therefore, characteristics of this particular implementation may not generalize to other tailored interface designs.

A limitation in this study is sample size. The baseline group included 22 participants, whereas the tailored interface group included 12 participants, of whom 11 contributed to the subjective and flight-path analysis and only 9 to the physiological analysis. A priori analysis in Section 6.1.1 targeted a large effect ($d = 0.80$), which required approximately 25 participants per group. The achieved sample size limited power to detect small to moderate effects, which helps explain why several comparisons produced meaningful directional differences and moderate effect sizes without reaching statistical significance. The seven-month gap between the two data collection phases introduces an additional cohort constraint, because the groups were tested at different points in the academic year under separate IRB protocols, and differences in training progression or seasonal flight activity between cohorts cannot be fully ruled out. Since the experiment used a between-subject design, individual differences in skill, scan strategy, and simulator familiarity were also uncontrolled. These constraints motivated the use of non-parametric tests and effect-size reporting, as justified in Section 6.3, and they are the primary reasons the results are presented as preliminary. Additionally, the pupil-dilation analysis relied on a pseudo-baseline constructed from the early task period rather than a true resting baseline, and pupil diameter remains sensitive to luminance, gaze angle, and headset fit, which constrains the physiological conclusions.

Another limitation concerns how the tailored interface was presented to participants. The interface ran on an external monitor that was brought into the Varjo XR-4 environment through video passthrough masking rather than rendered natively within the simulated cockpit. The masking process introduced blurring that may have reduced text sharpness and legibility relative to the native X-Plane instruments for some participants. Reduced legibility may have increased the visual effort required to read the interface, lengthened dwell times on the interface AOI, and attenuated the benefits the interface could otherwise provide. A dedicated external monitor viewed directly in a fixed base simulator would have preserved native display resolution. Alternatively, rendering the interface inside the X-Plane virtual cockpit itself would have presented it at the same fidelity as the surrounding instruments and better represented an integrated avionics implementation. Future evaluation should adopt one of these presentation methods so that measured effects reflect the information design of the interface rather than legibility of its rendering.

Finally, the evaluation used a single scenario, the Mammoth Yosemite RNAV (GPS) RWY 27 approach, which limits generalization to other approaches, terrain profiles, and operational contexts. Subjective measures were collected once, after the task, and therefore reflect retrospective impressions rather than moment-to-moment workload.

Despite these limitations, the convergence of directional findings across subjective, physiological, operational, and visual-attention measures supports the framework as a promising basis for future, better-powered evaluation. The findings are bounded to single-pilot, single-engine general aviation IFR operations. Participants were predominantly low-time pilots (mean flight hours experience of 200-235 hours) drawn from a collegiate flight program; the aircraft model was a Cessna 172. Future research should evaluate the framework with transport-category aircraft, multicrew operations, high flight hours pilots, and additional procedures, terrain profiles, and operational conditions. The framework may also transfer to other operational contexts including driving, surgery, and spaceflight. To apply the framework in each domain, researchers would identify context critical information from relevant operational and safety data, organize that information within the framework’s interface regions, validate the resulting design with representative users and evaluate its effect with physiological and subjective measures.

7.2 Future Work

Future research should evaluate the framework using a larger, more balanced sample of participants. The small, tailored-interface sample limited statistical power and made the results sensitive to individual differences. A larger sample with a wider range of pilots would support stronger statistical testing, more reliable effect-size estimates, and subgroup analysis based on pilot experience, training background, and simulator familiarity. Additionally, the present study compared different participants across baseline and tailored-interface conditions. Future studies should include a within-subjects design. A within-subjects design would allow each pilot to complete both conditions and serve as their own control. This design would reduce the influence of individual differences in skill, confidence, scan strategy, and flight experience.

Future work should continue refining the interface through pilot-centered usability testing. Pilots and flight instructors should specifically evaluate alert timing, map scale, color coding, text density, symbol placement, and information hierarchy, with a design goal of minimizing the information required for rapid comprehension and timely decision-making. Although I used mountainous operations for the purpose of validation, the interface should also apply to other operations such as high density-altitude departures, low-visibility approaches, missed approaches, emergency diversions, terrain-limited rotorcraft operations, and abnormal aircraft-performance scenarios. Each operation would require a different set of prioritized information elements while preserving the same framework logic.

A future system could adjust information priority based on flight phase, proximity to terrain, aircraft energy state, weather, or performance margin. However, the interface must remain transparent to the pilot. It should support pilot authority, avoid automation surprise, and reduce cognitive integration demands without creating a new monitoring burden. Later experiments should examine whether pilots make earlier go-around decisions, maintain safer terrain clearance, reduce exposure to unstable approaches, improve response time to performance limitations, or avoid continuing flight into unsafe conditions. These outcomes would provide stronger evidence that context-sensitive information presentation can improve operational safety.

8 References

Agramanisti Azdy, R., & Darnis, F. (2020). Use of Haversine Formula in Finding Distance Between Temporary Shelter and Waste End Processing Sites. *Journal of Physics: Conference Series*, *1500*(1), 012104. https://doi.org/10.1088/1742-6596/1500/1/012104

Alcantara, J. M. A., Plaza-Florido, A., Amaro-Gahete, F. J., Acosta, F. M., Migueles, J. H., Molina-Garcia, P., Sacha, J., Sanchez-Delgado, G., & Martinez-Tellez, B. (2020). Impact of Using Different Levels of Threshold-Based Artefact Correction on the Quantification of Heart Rate Variability in Three Independent Human Cohorts. *Journal of Clinical Medicine*, *9*(2), 325. https://doi.org/10.3390/jcm9020325

Andre, A. D., Wickens, C. D., & Moorman, L. (1991). Display Formatting Techniques for Improving Situation Awareness in the Aircraft Cockpit. *The International Journal of Aviation Psychology*, *1*(3), 205–218. https://doi.org/10.1207/s15327108ijap0103_2

Anton, G. E. (2022). Procedural Checklist/Situational Awareness. In R. Gilani & J. L. Mills (Eds.), *Vascular Complications of Surgery and Intervention* (pp. 149–160). Springer International Publishing. https://doi.org/10.1007/978-3-030-86713-3_9

Arthur Iii, J. J., Prinzel Iii, L. J., Williams, S. P., & Kramer, L. J. (2006). *Synthetic vision enhanced surface operations and flight procedures rehearsal tool* (J. G. Verly & J. J. Guell, Eds.; p. 62260I). https://doi.org/10.1117/12.666636

Arthur, J. (Trey) J., Prinzel, L. J., Bailey, R. E., Shelton, K. J., & Nicholas, S. N. (2018). Evaluation of head-worn and head-up displays for use in enhanced flight vision system operations. *Optical Engineering*, *58*(05), 1. https://doi.org/10.1117/1.OE.58.5.051802

Bachmann, S., Kancler, D. E., & Curtis, C. (2006). Supporting the Expert and Novice in a Single User Interface. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(5), 714–717. https://doi.org/10.1177/154193120605000521

Baker, S. P., & Lamb, M. W. (1989). Hazards of mountain flying: Crashes in the Colorado Rockies. *Aviation, Space, and Environmental Medicine*, *60*(6), 531–536.

Bandodkar, N., & Singh, R. (2014). Affect Infusion in a Computer Based Multitasking Environment: An Empirical Investigation. *AMCIS 2014 Proceedings 9*. https://aisel.aisnet.org/amcis2014/HumanComputerInteraction/GeneralPresentations/9

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine*, *68*(7), 569–579.

Berntson, G. G., Thomas Bigger, J., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., Nagaraja, H. N., Porges, S. W., Saul, J. P., Stone, P. H., & Van Der Molen, M. W. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, *34*(6), 623–648. https://doi.org/10.1111/j.1469-8986.1997.tb02140.x

Bhattacharya, S. (2025). Digital Innovations for Predicting Weather, Navigation and Charting. In P. K. Dutta, M. Ricciuti, I. Bogrekci, & S. Suseelan, *Digital Transformation in Aviation Industry Operations* (1st ed., pp. 23–44). Routledge. https://doi.org/10.4324/9781003494720-2

Bhattarai, A., Alarcon-Aneiva, L. J., & Fala, N. (2025). Enhancing pilot decision-making in mountainous terrain using UI system for CFIT prevention. *23rd International Symposium on Aviation Psychology*. International Symposium on Aviation Psychology.

Bhattarai, A., & Fala, N. (2025, May 29). *Analysis of controlled flight into terrain (CFIT) accidents post 2008 using complex network theory* [Poster Presentation]. 23rd International Symposium on Aviation Psychology.

Botega, L. C., Pereira, G. M. C., Junior, V. A. P., & De Oliveira, A. C. M. (2019). Assessing Situation Awareness on Fusion-Driven Emergency Management Systems. *2019 22th International Conference on Information Fusion (FUSION)*, 1–8. https://doi.org/10.23919/FUSION43075.2019.9011355

Botello, C., Jernigan, J., Stimson, D., Marquardt, S. W., Kancler, D. E., Curtis, C. K., Barthol, D. R., Burneka, C. M., & Whited, V. T. (2006). *Aircraft Maintenance Intuitive Troubleshooting (AMIT)* (Final AFRL-HE-WP-TR-2007-0006). Airforce Research Laboratory.

Braarud, P. Ø. (2021). Investigating the validity of subjective workload rating (NASA TLX) and subjective situation awareness rating (SART) for cognitively complex human–machine work. *International Journal of Industrial Ergonomics*, *86*, 103233. https://doi.org/10.1016/j.ergon.2021.103233

Breen, B. C. (1999). Controlled Flight Into Terrain and the enhanced Ground Proximity Warning system. *IEEE Aerospace and Electronic Systems Magazine*, *14*(1), 19–24. https://doi.org/10.1109/62.738350

Bud, M. J., Mengert, P., Ransom, S., & Stearns, M. D. (1997). *General aviation accidents, 1983-1994: Identification of factors related to controlled-flight-into-terrain (CFIT) accidents* (Final Report DOT-VNTSC-FAA-97-8). https://rosap.ntl.bts.gov/view/dot/5872

Burian, B. K., Mosier, K. L., Fischer, U. M., & Kochan, J. A. (2023). New teams on the flight deck: Humans and context-sensitive information automation. In *Human Factors in Aviation and Aerospace* (pp. 59–87). Elsevier. https://doi.org/10.1016/B978-0-12-420139-2.00008-3

Buysse, Y. (2025). *Stress recovery in virtual reality: The impact of presence on stress recovery in nature vs. Urban environments* [Project]. University of Skovde. https://www.diva-portal.org/smash/get/diva2:2079750/FULLTEXT01.pdf

Cahill, J., McDonald, N., Morrison, R., & Lynch, D. (2016). The operational validation of new cockpit technologies supporting all conditions operations: A case study.

Cognition, Technology & Work, *18*(3), 479–509. https://doi.org/10.1007/s10111-016-0380-4

Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods*, *41*(1), 113–117. https://doi.org/10.3758/BRM.41.1.113

Chen, O., Paas, F., & Sweller, J. (2023). A Cognitive Load Theory Approach to Defining and Measuring Task Complexity Through Element Interactivity. *Educational Psychology Review*, *35*(2), 63. https://doi.org/10.1007/s10648-023-09782-w

Chérif, L., Wood, V., Marois, A., Labonté, K., & Vachon, F. (2018). Multitasking in the military: Cognitive consequences and potential solutions. *Applied Cognitive Psychology*, *32*(4), 429–439. https://doi.org/10.1002/acp.3415

Cherven, K. (2015). *Mastering Gephi network visualization: Produce advanced network graphs in Gephi and gain valuable insights into your network datasets* (1st ed). Packt Publishing.

Chou, C.-C., Madhavan, D., & Funk, K. (1996). Studies of Cockpit Task Management Errors. *The International Journal of Aviation Psychology*, *6*(4), 307–320. https://doi.org/10.1207/s15327108ijap0604_1

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*, *51*(4), 661–703. https://doi.org/10.1137/070710111

Colvin, K., Funk, K., & Braune, R. (2009). Task Prioritization Factors: Two Part-Task Simulator Studies. *International Journal of Aviation Psychology*, *15*(4), 321–338. https://doi.org/https://doi.org/10.1207/s15327108ijap1504_2

Crump, C., Walenchok, S., Johnson, C., Pauszek, J., & Young, D. (2021). Stress, Operator Error, and Transportation Accidents. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 1453–1458. https://doi.org/10.1177/1071181321651171

Curtis, M. T., Jentsch, F., & Wise, J. A. (2010). Aviation Displays. In *Human Factors in Aviation* (pp. 439–478). Elsevier. https://doi.org/10.1016/B978-0-12-374518-7.00014-6

Damiba, A. (2024, October). *RASG-AFI Controlled Flight Into Terrain (CFIT) Workshop*. https://www.icao.int/WACAF/Documents/Meetings/2024/Workshop-CFIT/4-CFIT%20Workshop%20-%20AFPP%20PBN%20Regulatory%20approval%20and%20impact%20on%20CFIT.pdf

Davis, R. C., Wilt, D., Henion, J., Alter, K., & Snow, P. (2005). Affect of Predictive and Reactive Flying on Pilot Performance for Segmented Approaches. *24th Digital Avionics Systems Conference*, *2*, 13.B.4-1-13.B.4-7. https://doi.org/10.1109/DASC.2005.1563430

De Almeida, M. F., Soares, A. B. F., & Campos, F. A. D. (2026). Evaluation of Workload in Military Pilots During Formation Flight Instruction: Impacts of Function and

Experience. *Military Medicine*, *191*(1–2), e212–e218. https://doi.org/10.1093/milmed/usaf300

Degani, A., & Wiener, E. L. (1993). Cockpit Checklists: Concepts, Design, and Use. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *35*(2), 345–359. https://doi.org/10.1177/001872089303500209

Devlin, S. P., Moacdieh, N. M., Wickens, C. D., & Riggs, S. L. (2020). Transitions Between Low and High Levels of Mental Workload can Improve Multitasking Performance. *IISE Transactions on Occupational Ergonomics and Human Factors*, *8*(2), 72–87. https://doi.org/10.1080/24725838.2020.1770898

Dill, E. T., & Young, S. D. (2015, June 22). Analysis of Eye-Tracking Data with Regards to the Complexity of Flight Deck Information Automation and Management—Inattentional Blindness, System State Awareness, and EFB Usage. *15th AIAA Aviation Technology, Integration, and Operations Conference*. 15th AIAA Aviation Technology, Integration, and Operations Conference. https://doi.org/10.2514/6.2015-2901

Ding, L., Yuan, M., & Yu, R. (2025). Ergonomics Evaluation Modeling of Multi-modal Interaction Technology for Airplane Cockpit and Experimental Validation. In S. Jin, J. H. Kim, Y.-K. Kong, J. Park, & M. H. Yun (Eds.), *Proceedings of the 22nd Congress of the International Ergonomics Association, Volume 1* (Vol. 39, pp. 419–425). Springer Nature Singapore. https://doi.org/10.1007/978-981-95-0211-0_64

Dobson, J. E., & Campbell, J. S. (2014). The Flatness of U.S. States. *Geographical Review*, *104*(1), 1–9. https://doi.org/10.1111/j.1931-0846.2014.12001.x

DOD. (1984). *Legends for use in Aircrew Station and on Airborne Equipment* (MIL-STD-783D).

DOD. (1991). *Aircrew Station Alerting System* (MIL-STD-411E).

DOD. (1996). *Department of Defense Design Criteria Standard* (MIL-STD-1472E).

DOD. (2008). *Department of Defense Interface Standard- Common Warfighting Symbology*. https://worldwind.arc.nasa.gov/milstd2525c/Mil-STD-2525C.pdf

Doyon-Poulin, P., Ouellette, B., & Robert, J.-M. (2014). Effects of visual clutter on pilot workload, flight performance and gaze pattern. *Proceedings of the International Conference on Human-Computer Interaction in Aerospace*, 1–7. https://doi.org/10.1145/2669592.2669656

Draeger, J. (2009). Visuelle Ergonomie im Cockpit: Die Bedeutung der Instrumenteerkennbarkeit für die Flugsicherheit. *Der Ophthalmologe*, *106*(4), 370–374. https://doi.org/10.1007/s00347-008-1879-8

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Ericson, C. A. (2005). *Hazard Analysis Techniques for System Safety* (1st ed.). Wiley. https://doi.org/10.1002/0471739421

Erjavac, A. J., Iammartino, R., & Fossaceca, J. M. (2018). Evaluation of preconditions affecting symptomatic human error in general aviation and air carrier aviation accidents. *Reliability Engineering & System Safety*, *178*, 156–163. https://doi.org/10.1016/j.ress.2018.05.021

Etherington, T. J., Kramer, L. J., Smith-Velazquez, L., & De Haag, M. U. (2020). Simulation Study of Technology for Predicted Flight Deck Alerting of Energy. *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, 1–8. https://doi.org/10.1109/DASC50938.2020.9256670

FAA. (2003, April 1). *AC No: 61-134, General Aviation Controlled Flight Into Terrain Awareness*. Federal Aviation Administration. https://www.faa.gov/documentLibrary/media/Advisory_Circular/ac61-134.pdf

FAA. (2014, October 7). *Electronic Flight Displays: AC 25-11B* [Advisory Circular]. https://www.faa.gov/documentlibrary/media/advisory_circular/ac_25-11b.pdf

FAA. (2016). *Human Factors Design Standard* (HF-STD-001B). FAA.

Fala, N. (2022). An Analysis of Fixed-Wing Stall-Type Accidents in the United States. *Aerospace*, *9*(4), 178. https://doi.org/10.3390/aerospace9040178

Fay, M. P., & Proschan, M. A. (2010). Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics Surveys*, *4*(none). https://doi.org/10.1214/09-SS051

Felice, F. D., Petrillo, A., Carlomusto, A., & Romano, U. (2013). Modelling application for cognitive reliability and error analysis method. *International Journal of Engineering and Technology*, *5*, 4450–4464. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d5ab002447ef75b7e6613f35e890ed3ccaed4643

Frank, J. D., & Aaseng, G. B. (2016, September 13). Transitioning Autonomous Systems Technology Research to a Flight Software Environment. *AIAA SPACE 2016*. AIAA SPACE 2016. https://doi.org/10.2514/6.2016-5530

Fu, G., Yin, W.-T., Dong, J.-Y., Fan, D., & Zhu, C. J. (2013). Behavior-based accident causation: The “2-4” model and its safety implications in coal mines. *Meitan Xuebao/Journal of the China Coal Society*, *38*(7), 1123–1129. Scopus.

Gaetan, S., Dousset, E., Marqueste, T., Bringoux, L., Bourdin, C., Vercher, J.-L., & Besson, P. (2015). Cognitive Workload and Psychophysiological Parameters During Multitask Activity in Helicopter Pilots. *Aerospace Medicine and Human Performance*, *86*(12), 1052–1057. https://doi.org/10.3357/AMHP.4228.2015

Gajos, K. Z., Everitt, K., Tan, D. S., Czerwinski, M., & Weld, D. S. (2008). Predictability and accuracy in adaptive user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1271–1274. https://doi.org/10.1145/1357054.1357252

Gevrek, I., Ata, N., Gevrek, M. I., & Eroglu, C. (2026). Physiological Factors and Subjective Fatigue Monitoring in Helicopter Pilots During UH-60 Missions.

Aerospace Medicine and Human Performance, *97*(1), 20–25. https://doi.org/10.3357/AMHP.6749.2026

Godfroy-Cooper, M., Miller, J. D., Bachelder, E. N., & Szoboslay, Z. (2022). *OPERATOR STATE MONITORING FOR WORKLOAD PREDICTION AND MANAGEMENT*. Scopus. 48rd European Rotorcraft Forum, ERF 2022. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85151535591&partnerID=40&md5=56cc4b366d359d9135e39d3e11714c5e

Gumus, F., & Saylam, R. (2023). Prevention of Aviation Accidents with Prediction of Cognitive States. *2023 10th International Conference on Recent Advances in Air and Space Technologies (RAST)*, 1–4. https://doi.org/10.1109/RAST57548.2023.10197937

Hadar, B., Skrzypek, J. E., Wingfield, A., & Ben-David, B. M. (2016). Working Memory Load Affects Processing Time in Spoken Word Recognition: Evidence from Eye-Movements. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00221

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hart, S. G., & Staveland. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (Vol. 52, pp. 139–183). Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9

Hedges, L. V. (2008). What Are Effect Sizes and Why Do We Need Them? *Child Development Perspectives*, *2*(3), 167–171. https://doi.org/10.1111/j.1750-8606.2008.00060.x

Heinrich, H. W. (1941). *Industrial Accident Prevention*. McGrawl Hill Book Company. https://archive.org/details/dli.ernet.14601/page/n5/mode/2up

Henneberry, D., Ziakkas, D., Bokas, E., & Pechlivanis, K. (2026). *Cultural Dynamics in Next-Generation Cockpit Design: Integrating Human Factors, Inclusivity, and System Resilience in Transportation Aviation*. 9th International Conference on Human Intelligent Systems Integration (IHSI 2026): Disruptive and Innovative Technologies. https://doi.org/10.54941/ahfe1007086

Himi, S. A., Bühner, M., Schwaighofer, M., Klapetek, A., & Hilbert, S. (2019). Multitasking behavior and its related constructs: Executive functions, working memory capacity, relational integration, and divided attention. *Cognition*, *189*, 275–298. https://doi.org/10.1016/j.cognition.2019.04.010

Hofer, N., Miller, L., & Abdulrahim, M. (2024, January 8). Detecting Pilot Task Saturation. *AIAA SCITECH 2024 Forum*. AIAA SCITECH 2024 Forum. https://doi.org/10.2514/6.2024-2476

Holley, S., & Miller, M. (2022). Effects of Cognitive Loading on Pilots and Air Traffic Controller Performance: Implications for Neural Dynamics and Cognitive Flow.

Proceedings of the Human Factors and Ergonomics Society Annual Meeting, *66*(1), 2256–2260. https://doi.org/10.1177/1071181322661544

Honeywell. (2000). *EGPWS MK V and MK VII Enhanced Ground Proximity Warning system pilots guide: D*. Honeywell.

Hsueh, C.-H., & Chen, C.-H. (2025). A Study on the Visual Cue Designs in Rich Notification User Interface. In H. Mori & Y. Asahi (Eds.), *Human Interface and the Management of Information* (Vol. 15773, pp. 90–101). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-93819-1_7

Huttner, J.-P., Tews, L., Laudien, T., & Friedrich, M. (2024, July 29). Bridging the Gap: Adaptive HMI Innovations for Next-Gen Fighter Cockpits. *AIAA AVIATION FORUM AND ASCEND 2024*. AIAA AVIATION FORUM AND ASCEND 2024. https://doi.org/10.2514/6.2024-4215

IATA. (n.d.). *Interactive Safety Report- International Air Transport Association*. Retrieved https://www.iata.org/en/publications/safety-report/interactive-safety-report/

ICAO. (1991). *STOLport Manual DOC 9150-AN/899*. International Civil Aviation Organization.

Janetzko, D., & Kacem, B. (2024). What Do You Need? Information Requirements and Task Analysis of (Future) Advanced Air Mobility Pilots in the Emergency Medical Service. *Aerospace*, *11*(3), 197. https://doi.org/10.3390/aerospace11030197

Johnson, J. S., Simmering, V. R., & Buss, A. T. (2014). Beyond slots and resources: Grounding cognitive concepts in neural dynamics. *Attention, Perception, & Psychophysics*, *76*(6), 1630–1654. https://doi.org/10.3758/s13414-013-0596-9

Johnson-Laird, P. N. (2012). Inference with Mental Models. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (1st ed., pp. 134–154). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0009

Jordan, J., Wagner, J., Manthey, D. E., Wolff, M., Santen, S., & Cico, S. J. (2020). Optimizing Lectures From a Cognitive Load Perspective. *AEM Education and Training*, *4*(3), 306–312. https://doi.org/10.1002/aet2.10389

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, *5*(2), 113–153. https://doi.org/10.1080/1463922021000054335

Kalagher, H., De Voogt, A., & Boulter, C. (2021). Situational Awareness and General Aviation Accidents: An Analysis of 94 US Accident Reports. *Aviation Psychology and Applied Human Factors*, *11*(2), 112–117. https://doi.org/10.1027/2192-0923/a000207

Katsakiori, P., Sakellaropoulos, G., & Manatakis, E. (2009). Towards an evaluation of accident investigation methods in terms of their alignment with accident causation models. *Safety Science*, *47*(7), 1007–1015. https://doi.org/10.1016/j.ssci.2008.11.002

Kelly, D., & Efthymiou, M. (2019). An analysis of human factors in fifty controlled flight into terrain aviation accidents from 2007 to 2017. *Journal of Safety Research*, *69*, 155–165. https://doi.org/10.1016/j.jsr.2019.03.009

Kelly, J. R., & Loving, T. J. (2004). Time pressure and group performance: Exploring underlying processes in the Attentional Focus Model. *Journal of Experimental Social Psychology*, *40*(2), 185–198. https://doi.org/10.1016/S0022-1031(03)00094-5

Khatwa, R., & Roelen, A. L. C. (1999). *An Analysis of Controlled Flight into Terrain (CFIT) Accidents of Commercial Operators 1988 through 1994* (Flight Safety Digest Vol 17, pp. 163–184).

Kim, S. H., Lee, J. H., Park, D., Lee, Y., & Yun, M. H. (2019). Effects of Visual Complexity Levels and Information Decluttering Methods for Cockpit Displays on Human Search Performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 96–100. https://doi.org/10.1177/1071181319631057

Kobayashi, H., Park, B.-J., & Miyazaki, Y. (2012). Normative references of heart rate variability and salivary alpha-amylase in a healthy young male population. *Journal of Physiological Anthropology*, *31*(1), 9. https://doi.org/10.1186/1880-6805-31-9

Krause, E., Vollmer, M., Wittfeld, K., Weihs, A., Frenzel, S., Dörr, M., Kaderali, L., Felix, S. B., Stubbe, B., Ewert, R., Völzke, H., & Grabe, H. J. (2023). Evaluating heart rate variability with 10 second multichannel electrocardiograms in a large population-based sample. *Frontiers in Cardiovascular Medicine*, *10*, 1144191. https://doi.org/10.3389/fcvm.2023.1144191

Ladurini, A., Crate, S., Mitchell, M., Sanchez, S., Eble, R., & Brockhagen, T. (2024). Reducing Workload: A Double-Edged Sword. *2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference (DASC)*, 1–7. https://doi.org/10.1109/DASC62030.2024.10749018

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Lauren, D., Arnaud, B., Paul, K., Hervé, G., & Olivier, M. (2018). Using Machine Learning Algorithms to Develop Adaptive Man–Machine Interfaces. In *Neuroergonomics* (pp. 237–238). Elsevier. https://doi.org/10.1016/B978-0-12-811926-6.00053-1

Lazaro, M. J., Kang, Y., Yun, M. H., & Kim, S. (2021). The Effects of Visual Complexity and Decluttering Methods on Visual Search and Target Detection in Cockpit Displays. *International Journal of Human–Computer Interaction*, *37*(7), 588–600. https://doi.org/10.1080/10447318.2021.1890491

Lazaro, M. J., Kim, S., Kang, Y., & Hwan Yun, M. (2021). Visual Search and Decluttering in Tactical Situation Displays: A Computational Modeling Approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 1425–1431. https://doi.org/10.1177/1071181321651154

Lee, J. (2012). Modeling terrain awareness and warning systems for airspace and procedure design. *2012 IEEE/AIAA 31st Digital Avionics Systems Conference (DASC)*, 2B2-1-2B2-9. https://doi.org/10.1109/DASC.2012.6382283

Letsu-Dake, E., Pepitone, D., Ball, J., & Burgin, R. (2014). Recommendations for managing complexity in electronic chart information displays. *2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC)*, 4B1-1-4B1-14. https://doi.org/10.1109/DASC.2014.6979467

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science*, *42*(4), 237–270. https://doi.org/10.1016/S0925-7535(03)00047-X

Li, K., & Wang, S. (2018). A network accident causation model for monitoring railway safety. *Safety Science*, *109*, 398–402. https://doi.org/10.1016/j.ssci.2018.06.008

Li, K., Zhai, J., Liu, Q., Dai, S., & Wang, Q. (2026). Scan Pattern and Heart Rate Variability During Simulated Flight Task. In S. Long, B. S. Dhillon, & L. Ye (Eds.), *Man-Machine-Environment System Engineering* (Vol. 1476, pp. 84–93). Springer Nature Singapore. https://doi.org/10.1007/978-981-95-1904-0_12

Li, S., Chen, G., Zhou, L., Zhao, Y., Xu, Q., & Zhao, J. (2025). Complex network analysis for accident causes modelling to enhance process safety in chemical enterprises. *The Canadian Journal of Chemical Engineering*, *103*(2), 679–696. https://doi.org/10.1002/cjce.25443

Li, W.-C., Horn, A., Sun, Z., Zhang, J., & Braithwaite, G. (2020). Augmented visualization cues on primary flight display facilitating pilot's monitoring performance. *International Journal of Human-Computer Studies*, *135*, 102377. https://doi.org/10.1016/j.ijhcs.2019.102377

Li, Y., Dikmen, M., Hussein, T. G., Wang, Y., & Burns, C. (2018). To Cross or Not to Cross: Urgency-Based External Warning Displays on Autonomous Vehicles to Improve Pedestrian Crossing Safety. *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 188–197. https://doi.org/10.1145/3239060.3239082

Lim, Y., Gardi, A., Sabatini, R., Ramasamy, S., Kistan, T., Ezer, N., Vince, J., & Bolia, R. (2018). Avionics Human-Machine Interfaces and Interactions for Manned and Unmanned Aircraft. *Progress in Aerospace Sciences*, *102*, 1–46. https://doi.org/10.1016/j.paerosci.2018.05.002

Lin, L., Wang, Q., & Sadek, A. W. (2014). Data Mining and Complex Network Algorithms for Traffic Accident Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *2460*(1), 128–136. https://doi.org/10.3141/2460-14

Liu, D., Guarino, S. L., Roth, E., Harper, K., & Vincenzi, D. (2012). Effect of Novel Adaptive Displays on Pilot Performance and Workload. *The International Journal of Aviation Psychology*, *22*(3), 242–265. https://doi.org/10.1080/10508414.2012.691040

Liu, H., Dong, D., & Meng, H. (2018). A Compliance Method for the Design and Airworthiness Certification of Civil Aircraft Flight Deck Human Factor. In C.

Stephanidis (Ed.), *HCI International 2018 – Posters' Extended Abstracts* (Vol. 850, pp. 79–84). Springer International Publishing. https://doi.org/10.1007/978-3-319-92270-6_11

Liu, Q., Wang, Y., Bai, Y., Yu, M., Cao, Z., Yu, X., & Ding, L. (2023). Development of a quantitative measurement on visual clutter in see through display. *Frontiers in Neuroscience*, *17*, 1138225. https://doi.org/10.3389/fnins.2023.1138225

Liu, S., Wadeson, A., Kim, N. Y., & Nam, C. S. (2016). Effects of Working Memory Capacity, Task Switching, and Task Difficulty on Multitasking Performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 502–506. https://doi.org/10.1177/1541931213601114

Loft, S., Tatasciore, M., & Visser, T. (2023). Managing workload, performance, and situation awareness in aviation systems. In *Human Factors in Aviation and Aerospace* (pp. 171–197). Elsevier. https://doi.org/10.1016/B978-0-12-420139-2.00018-6

Logie, R. H., Law, A., Trawley, S., & Nissan, J. (2010a). Multitasking, working memory and remembering intentions. *PSYCHOL BELG*, *50*(3–4), 309. https://doi.org/10.5334/pb-50-3-4-309

Logie, R. H., Law, A., Trawley, S., & Nissan, J. (2010b). Multitasking, working memory and remembering intentions. *Psychologica Belgica*, *50*(3–4), 309. https://doi.org/10.5334/pb-50-3-4-309

López-Contreras, E. (2025). [DC] IMMERCOG: Impact of Extreme Environments on Remote Operations in Critical Systems. *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 1584–1585. https://doi.org/10.1109/VRW66409.2025.00437

Lutnyk, L., Kim, K. G., Sarbach, A., Kiefer, P., Häusler, R., & Raubal, M. (2025). Context-Sensitive Augmented Reality Assistance in the Cockpit. *International Journal of Human–Computer Interaction*, *41*(17), 11149–11168. https://doi.org/10.1080/10447318.2024.2440976

Madhavan, P., & Lacson, F. C. (2006). Psychological factors affecting pilots' decisions to navigate in deteriorating weather. *North American Journal of Psychology*, *8*(1), 47–62. Scopus.

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491

Martin, K., McLeod, E., Périard, J., Rattray, B., Keegan, R., & Pyne, D. B. (2019). The Impact of Environmental Stress on Cognitive Performance: A Systematic Review. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *61*(8), 1205–1246. https://doi.org/10.1177/0018720819839817

Martin, W. L., Murray, P. S., & Bates, P. R. (2011). *REDUCING PATHOLOGICAL STRESS EFFECTS AND INCREASING PILOT PERFORMANCE DURING UNEXPECTED IN-FLIGHT EVENTS.* 381–386. https://corescholar.libraries.wright.edu/isap_2011/51

McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Meier, E. D. (2021). Eye tracking study for improving tactile feedback in an aviation task. *Gis.Science - Die Zeitschrift Fur Geoinformatik*, *2021*(2), 57–65. Scopus.

Mejdal, S., McCauley, M. E., & Beringer, D. B. (2001). *Human Factors Design Guidelines for Multifunction Displays* (Final Report DOT/FAA/AM-01/17). FAA.

Meng, B., & Lu, N. (2022). A Hybrid Model Integrating HFACS and BN for Analyzing Human Factors in CFIT Accidents. *Aerospace*, *9*(11), 711. https://doi.org/10.3390/aerospace9110711

Menig, A., Becker, N., & Kleudgen, J. P. (2025). Investigating the role of an AI-based assistance system for the decision-making process of pilots– an interview study. *Transportation Research Procedia*, *88*, 167–175. https://doi.org/10.1016/j.trpro.2025.05.021

Michailovs, S., Pond, S., Schmitt, M., Irons, J., Stoker, M., Visser, T. A. W., Huf, S., & Loft, S. (2023). The Impact of Information Integration in a Simulation of Future Submarine Command and Control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *65*(7), 1473–1490. https://doi.org/10.1177/00187208211045872

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Mishra, M., Sidoti, D., Avvari, G. V., Mannaru, P., Ayala, D. F. M., Pattipati, K. R., & Kleinman, D. L. (2017). A Context-Driven Framework for Proactive Decision Support With Applications. *IEEE Access*, *5*, 12475–12495. https://doi.org/10.1109/ACCESS.2017.2707091

Moacdieh, N. M., & Sarter, N. (2017). The Effects of Data Density, Display Organization, and Stress on Search Performance: An Eye Tracking Study of Clutter. *IEEE Transactions on Human-Machine Systems*, *47*(6), 886–895. https://doi.org/10.1109/THMS.2017.2717899

MoCTCA. (n.d.). *Accident Investigation of 9N-AET, DHC-6/300* [Final Report]. Retrieved https://www.tourism.gov.np//files/publication_files/346.pdf

Moehlenbrink, C., Peinecke, N., Papenfuß, A., Manske, P., & Wies, M. (2011). *Investigating attentional tunneling through a flexible experimentation environment and eye tracking* (J. T. Thomas, D. D. Desjardins, J. J. Güell, & K. L. Bernier, Eds.; p. 80420Z). https://doi.org/10.1117/12.886714

Moens, L. L., Lydon, S., Cucurachi, S., O'Connor, P., Sauter, T. C., Töndury, G.-A., & Manser, T. (2026). Measuring Situation Awareness: A Meta-Review Across

Domains. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 00187208251412110. https://doi.org/10.1177/00187208251412110

Müller, M. (2004). [Risk and risk management in aviation]. *Zeitschrift Fur Arztliche Fortbildung Und Qualitatssicherung*, *98*(7), 559–565.

Müller, S., & Patel, H. R. H. (2012). Lessons Learned from the Aviation Industry: Surgical Checklists. In H. R. H. Patel & J. V. Joseph (Eds.), *Simulation Training in Laparoscopy and Robotic Surgery* (pp. 1–6). Springer London. https://doi.org/10.1007/978-1-4471-2930-1_1

Niermann, C. A. (2015). Can spatial audio support pilots? 3D-audio for future pilot-assistance systems. *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, 3C51-3C57. https://doi.org/10.1109/DASC.2015.7311401

Norman, D. (2014). Some Observation on Mental Model. In D. Gentner (Ed.), *Mental Models* (0 ed., pp. 8–19). Psychology Press. https://doi.org/10.4324/9781315802725

NTSB. (n.d.). *NTSB Aviation Accident Database* [Dataset]. Retrieved https://data.ntsb.gov/avdata

Oniscenko, B., Socha, V., Hanakova, L., Tlapak, J., & Matowicki, M. (2024). Impact of mild hypoxia on pilots' performance and physiological response: A systematic review and experimental study. *International Journal of Industrial Ergonomics*, *104*, 103650. https://doi.org/10.1016/j.ergon.2024.103650

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140–160. https://doi.org/10.1518/155534308X284417

Peltola, M. A. (2012). Role of editing of R–R intervals in the analysis of heart rate variability. *Frontiers in Physiology*, *3*. https://doi.org/10.3389/fphys.2012.00148

Pereira, T., Almeida, P. R., Cunha, J. P. S., & Aguiar, A. (2017). Heart rate variability metrics for fine-grained stress level assessment. *Computer Methods and Programs in Biomedicine*, *148*, 71–80. https://doi.org/10.1016/j.cmpb.2017.06.018

Pham, T., Lau, Z. J., Chen, S. H. A., & Makowski, D. (2021). Heart Rate Variability in Psychology: A Review of HRV Indices and an Analysis Tutorial. *Sensors*, *21*(12), 3998. https://doi.org/10.3390/s21123998

Pink, A., & Newton, P. M. (2020). Decorative animations impair recall and are a source of extraneous cognitive load. *Advances in Physiology Education*, *44*(3), 376–382. https://doi.org/10.1152/advan.00102.2019

Plews, D. J., Laursen, P. B., Kilding, A. E., & Buchheit, M. (2012). Heart rate variability in elite triathletes, is variation in variability the key to effective training? A case comparison. *European Journal of Applied Physiology*, *112*(11), 3729–3741. https://doi.org/10.1007/s00421-012-2354-4

Provan, D. J., Woods, D. D., Dekker, S. W. A., & Rae, A. J. (2020). Safety II professionals: How resilience engineering can transform safety practice. *Reliability Engineering & System Safety*, *195*, 106740. https://doi.org/10.1016/j.ress.2019.106740

Ramos, M. A., Sankaran, K., Guarro, S., Mosleh, A., Ramezani, R., & Arjounilla, A. (2023). The need for and conceptual design of an AI model-based Integrated Flight Advisory System. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *237*(2), 485–507. https://doi.org/10.1177/1748006X221083379

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. *Safety Science*, *27*(2–3), 183–213. https://doi.org/10.1016/S0925-7535(97)00052-0

Reason, J. (2000). Human error: Models and management. *BMJ (Clinical Research Ed.)*, *320*(7237), 768–770. https://doi.org/10.1136/bmj.320.7237.768

Regmi, R. P., Kitada, T., Dudhia, J., & Maharjan, S. (2017). Large-Scale Gravity Current over the Middle Hills of the Nepal Himalaya: Implications for Aircraft Accidents. *Journal of Applied Meteorology and Climatology*, *56*(2), 371–390. https://doi.org/10.1175/JAMC-D-16-0073.1

Reid, N. (2009). The concept of working memory: Introduction to the Special Issue. *Research in Science & Technological Education*, *27*(2), 131–137. https://doi.org/10.1080/02635140902853608

Richards, D., & Lamb, P. (2016, June 13). Functional Symbology—Evaluation of task-specific Head-Up Display information for use on a commercial flight deck. *AIAA Modeling and Simulation Technologies Conference*. AIAA Modeling and Simulation Technologies Conference. https://doi.org/10.2514/6.2016-3374

Rodriguez-Paras, C., McKenzie, J. T., Choterungruengkorn, P., & Ferris, T. K. (2021). Severity-Mapped Vibrotactile Cues to Support Interruption Management with Weather Messaging in the General Aviation Cockpit. *Atmosphere*, *12*(3), 341. https://doi.org/10.3390/atmos12030341

Rothrock, L., Koubek, R., Fuchs, F., Haas, M., & Salvendy, G. (2002). Review and reappraisal of adaptive interfaces: Toward biologically inspired paradigms. *Theoretical Issues in Ergonomics Science*, *3*(1), 47–84. https://doi.org/10.1080/14639220110110342

Rudi, D., Kiefer, P., Giannopoulos, I., & Raubal, M. (2020). Gaze-based interactions in the cockpit of the future: A survey. *Journal on Multimodal User Interfaces*, *14*(1), 25–48. https://doi.org/10.1007/s12193-019-00309-8

Ruijters, E., & Stoelinga, M. (2015). Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools. *Computer Science Review*, *15–16*, 29–62. https://doi.org/10.1016/j.cosrev.2015.03.001

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, *35*(11), 991–1009. https://doi.org/10.1016/j.ergon.2005.04.005

Saffell, T. N., Alexander, A. L., Carlin, A. S., Chang, A. C., & Schurr, N. (2011). *An Integrated Alerting and Notification System Utilizing Stages of Automation and Uncertainty Visualization*. 197–202.

Salvador, M., Yacout, S., & AboElHassan, A. (2022). Using Big Data and Machine Learning to Improve Aircraft Reliability and Safety. *2022 Annual Reliability and Maintainability Symposium (RAMS)*, 1–9. https://doi.org/10.1109/RAMS51457.2022.9894015

Salvendy, G., & Karwowski, W. (Eds.). (2021). *Handbook of human factors and ergonomics* (Fifth edition). John Wiley & Sons, Inc.

Sarter, N. B., & Woods, D. D. (1997). Team Play with a Powerful and Independent Agent: Operational Experiences and Automation Surprises on the Airbus A-320. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(4), 553–569. https://doi.org/10.1518/001872097778667997

Scarr, J., Cockburn, A., Gutwin, C., & Malacria, S. (2013). Testing the robustness and performance of spatially consistent interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3139–3148. https://doi.org/10.1145/2470654.2466430

Schlimm, K. A. (2015, June 22). A Model for Situational Awareness in Aircraft Upset Prevention and Recovery. *15th AIAA Aviation Technology, Integration, and Operations Conference*. 15th AIAA Aviation Technology, Integration, and Operations Conference. https://doi.org/10.2514/6.2015-2900

Seidler, K. S., & Wickens, C. D. (1992). Distance and organization in multifunction displays. *Human Factors*, *34*(5), 555–569.

Senol, M. B. (2020). A new optimization model for design of traditional cockpit interfaces. *Aircraft Engineering and Aerospace Technology*, *92*(3), 404–417. https://doi.org/10.1108/AEAT-04-2019-0068

Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2016). Object selection costs in visual working memory: A diffusion model analysis of the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1673–1693. https://doi.org/10.1037/a0040213

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*, 258. https://doi.org/10.3389/fpubh.2017.00258

Shaffer, F., Meehan, Z. M., & Zerr, C. L. (2020). A Critical Review of Ultra-Short-Term Heart Rate Variability Norms Research. *Frontiers in Neuroscience*, *14*, 594880. https://doi.org/10.3389/fnins.2020.594880

Shahriari, M., & Aydin, M. E. (2018). Aviation Accident Analysis: A Case Study. In P. Arezes (Ed.), *Advances in Safety Management and Human Factors* (Vol. 604, pp. 77–85). Springer International Publishing. https://doi.org/10.1007/978-3-319-60525-8_9

Sherman, B. E., & Turk-Browne, N. B. (2024). Attention and Memory. In M. J. Kahana & A. D. Wagner (Eds.), *The Oxford Handbook of Human Memory, Two Volume Pack* (1st ed., pp. 587–613). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190917982.013.21

Sherry, L., Polson, P., & Feary, M. (2002). *Designing User-Interfaces for the Cockpit: Five Common Design Errors and How to Avoid Them*. 2002-01–2968. https://doi.org/10.4271/2002-01-2968

Shupp, L., Andrews, C., Dickey-Kurdziolek, M., Yost, B., & North, C. (2009). Shaping the Display of the Future: The Effects of Display Size and Curvature on User Performance and Insights. *Human–Computer Interaction*, *24*(1–2), 230–272. https://doi.org/10.1080/07370020902739429

Singer, G., & Dekker, S. W. A. (2000). Pilot Performance During Multiple Failures: An Empirical Study of Different Warning Systems. *Transportation Human Factors*, *2*(1), 63–76. https://doi.org/10.1207/STHF0201_8

Soares, A. B. F., Almeida, M. F. D., Franchini, E., Ferreira, F. G., & Campos, F. A. D. (2025). Heart Rate Variability in Military Pilots During Flight: A Scoping Review. *Military Medicine*, *190*(3–4), e515–e522. https://doi.org/10.1093/milmed/usae390

Sohn, Y. W., & Doane, S. M. (2004). Memory Processes of Flight Situation Awareness: Interactive Roles of Working Memory Capacity, Long-Term Working Memory, and Expertise. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(3), 461–475. https://doi.org/10.1518/hfes.46.3.461.50392

Souvestre, P. A., Landrock, C. K., & Blaber, A. P. (2008). New paradigm for understanding in-flight decision making errors: A neurophysiological model leveraging human factors. *Hippokratia*, *12*(SUPPL. 1), 78–83. Scopus.

Spector, R. H. (1990). The Pupils. In H. K. Walker, W. D. Hall, & J. W. Hurst (Eds.), *Clinical Methods: The History, Physical, and Laboratory Examinations* (3rd ed.). Butterworths. http://www.ncbi.nlm.nih.gov/books/NBK381/

Stanton, N. A., Harris, D., Salmon, P. M., Demagalski, J. M., Marshall, A., Young, M. S., Dekker, S. W. A., & Waldmann, T. (2006). Predicting design induced pilot error using HET (human error template) – A new formal human error identification method for flight decks. *The Aeronautical Journal*, *110*(1104), 107–115. https://doi.org/10.1017/S0001924000001056

Stasch, S.-M., & Mack, W. (2024). Diagnosing Cognitive Control with Eye-Tracking Metrics in a Multitasking Environment. In D. Harris & W.-C. Li (Eds.), *Engineering Psychology and Cognitive Ergonomics* (Vol. 14692, pp. 89–102). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-60728-8_8

Sumwalt III, R. L. (2004). *Enhancing flight-crew monitoring skills can increase corporate aviation safety*. *49*, 197–203. Scopus.

https://www.scopus.com/inward/record.uri?eid=2-s2.0-8344276857&partnerID=40&md5=158f60dc8ff0a4c6ba7075e8f13d0788

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Taheri Gorji, H., Wilson, N., VanBree, J., Hoffmann, B., Petros, T., & Tavakolian, K. (2023). Using machine learning methods and EEG to discriminate aircraft pilot cognitive workload during flight. *Scientific Reports*, *13*(1), 2507. https://doi.org/10.1038/s41598-023-29647-0

Tamilselvan, G., Dinler, N., & Belt, S. (2023). *Performance Vs. Workload Matrix of Primary Flight Training: Exploratory Study*. 14th International Conference on Applied Human Factors and Ergonomics (AHFE 2023). https://doi.org/10.54941/ahfe1003166

Tang, H., Lee, B. G., Towey, D., & Pike, M. (2024). The Impact of Various Cockpit Display Interfaces on Novice Pilots' Mental Workload and Situational Awareness: A Comparative Study. *Sensors*, *24*(9), 2835. https://doi.org/10.3390/s24092835

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV – Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*(1), 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024

Taylor, R. M. (2017). Situational Awareness Rating Technique (Sart): The Development of a Tool for Aircrew Systems Design. In E. Salas & A. S. Dietz, *Situational Awareness* (1st ed., pp. 111–128). Routledge. https://doi.org/10.4324/9781315087924-8

Tian, F., Tian, Z., & Tian, Y. (2022). Research on Prediction and Quantitative Analysis of Controlled Flight into Terrain Impact Based on QAR Over-limit Events. *2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, 155–163. https://doi.org/10.1109/IMCEC55388.2022.10020144

Torres, S., Dehn, J., McKay, E., Paglione, M. M., & Schnitzer, B. S. (2015). En-Route Automation Modernization (ERAM) trajectory model evolution to support trajectory-based operations (TBO). *2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*, 1A6-1-1A6-15. https://doi.org/10.1109/DASC.2015.7311332

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Troller, M., Scaramuzza, M., Wipf, H., Nyffenegger, M., & Leibundgut, H. (2016). *Flight Performance Investigations of Enhanced Rotorcraft Operations in Mountainous Areas � Towards a More Ambitious RNP Performance*. 3660–3668. https://doi.org/10.33012/2016.14813

Troyer, G., Hagl, M., & Jakobi, J. (2025). *Mayday, Mayday! - Is Heart Rate Variability a Suitable Objective Indicator to Detect Pilot's Increased Mental Workload in Emergency Situations?* 16th International Conference on Applied Human Factors and Ergonomics (AHFE 2025). https://doi.org/10.54941/ahfe1006490

Tufano, F., Bahadure, S. W., Tufo, M., Novella, L., Fiengo, G., & Santini, S. (2023). An Optimization Framework for Information Management in Adaptive Automotive Human–Machine Interfaces. *Applied Sciences*, *13*(19), 10687. https://doi.org/10.3390/app131910687

U.S. Army Aviation Development Directorate, Bachelder, E., NASA Ames Research Center, Godfroy-Cooper, M., Miller, J. D., Haifa University, Kahana, A., Mafat HFE, & Rottem-Hovev, M. (2018). Multimodal Cueing and Assessment of Pilot Engagement During Low Level Flight. *Proceedings of the Vertical Flight Society 74th Annual Forum*, 1–19. https://doi.org/10.4050/F-0074-2018-12747

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Vanderhaegen, F. (2015). Erik Hollnagel: Safety-I and Safety-II, the past and future of safety management. *Cognition, Technology & Work*, *17*(3), 461–464. https://doi.org/10.1007/s10111-015-0345-z

Vidal, F., Meckler, C., & Hasbroucq, T. (2015). Basics for sensorimotor information processing: Some implications for learning. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00033

Virtanen, K., Mansikka, H., Kontio, H., & Harris, D. (2022). Weight watchers: NASA-TLX weights revisited. *Theoretical Issues in Ergonomics Science*, *23*(6), 725–748. https://doi.org/10.1080/1463922X.2021.2000667

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, H., Jiang, N., Pan, T., Si, H., Li, Y., & Zou, W. (2020). Cognitive Load Identification of Pilots Based on Physiological-Psychological Characteristics in Complex Environments. *Journal of Advanced Transportation*, *2020*, 1–16. https://doi.org/10.1155/2020/5640784

Wang, P., Fang, W., Guo, B., & Bao, H. (2018). Apply Petri Nets to Human Performance and Workload Prediction Under Multitask. In I. L. Nunes (Ed.), *Advances in Human Factors and Systems Interaction* (Vol. 592, pp. 395–405). Springer International Publishing. https://doi.org/10.1007/978-3-319-60366-7_37

Wehler, D., Jelinek, H. F., Gronau, A., Wessel, N., Kraemer, J. F., Krones, R., & Penzel, T. (2021). Reliability of heart-rate-variability features derived from ultra-short ECG recordings and their validity in the assessment of cardiac autonomic neuropathy. *Biomedical Signal Processing and Control*, *68*, 102651. https://doi.org/10.1016/j.bspc.2021.102651

Wei, Z., Zhuang, D., Bai, J., & Yang, K. (2013). *A new method for the analysis of design-induced pilot error on the flight decks*. 221–256. Scopus.

https://www.scopus.com/inward/record.uri?eid=2-s2.0-84896479881&partnerID=40&md5=c336eef166eea4156e9ef6950a9361e3

Wen, S., Middleton, M., Ping, S., Chawla, N. N., Wu, G., Feest, B. S., Nadri, C., Liu, Y., Kaber, D., Zahabi, M., McMahan, R. P., Castelo, S., McKendrick, R., Qian, J., & Silva, C. T. (2025). AdaptiveCoPilot: Design and Testing of a NeuroAdaptive LLM Cockpit Guidance System in both Novice and Expert Pilots. *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, 656–666. https://doi.org/10.1109/VR59515.2025.00088

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Wickens, C. D., & Carswell, C. M. (1995). The Proximity Compatibility Principle: Its Psychological Foundation and Relevance to Display Design. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(3), 473–494. https://doi.org/10.1518/001872095779049408

Wickens, C. D., & Carswell, C. M. (2021). INFORMATION PROCESSING. In G. Salvendy & W. Karwowski (Eds.), *HANDBOOK OF HUMAN FACTORS AND ERGONOMICS* (1st ed., pp. 114–158). Wiley. https://doi.org/10.1002/9781119636113.ch5

Wickens, C. D., Vidulich, M. A., & Tsang, P. S. (2023). Information processing in aviation. In *Human Factors in Aviation and Aerospace* (pp. 89–139). Elsevier. https://doi.org/10.1016/B978-0-12-420139-2.00002-2

Wong, D. K. Y., Pitfield, D. E., Caves, R. E., & Appleyard, A. J. (2006). Quantifying and characterising aviation accident risk factors. *Journal of Air Transport Management*, *12*(6), 352–357. https://doi.org/10.1016/j.jairtraman.2006.09.002

Woods, D. D. (2015). Four concepts for resilience and the implications for the future of resilience engineering. *Reliability Engineering & System Safety*, *141*, 5–9. https://doi.org/10.1016/j.ress.2015.03.018

Wu, C. M., Schulz, E., Pleskac, T. J., & Speekenbrink, M. (2022). Time pressure changes how people explore and respond to uncertainty. *Scientific Reports*, *12*(1), 4122. https://doi.org/10.1038/s41598-022-07901-1

Xenos, M., Mallas, A., & Minas, D. (2024). Using Eye-Tracking for Adaptive Human-Machine Interfaces for Pilots: A Literature Review and Sample Cases. *Journal of Physics: Conference Series*, *2716*(1), 012072. https://doi.org/10.1088/1742-6596/2716/1/012072

Xiao, X., Tong, J., Chen, P., Sun, J., Sui, Z., Liang, J., Zhao, H., Zhao, J., & Wang, H. (2025). *InSight-R: A Framework for Risk-informed Human Failure Event Identification and Interface-Induced Risk Assessment Driven by AutoGraph* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2507.00066

Xiao, X., Wanyan, X., & Zhuang, D. (2015). Mental workload prediction based on attentional resource allocation and information processing. *Bio-Medical Materials and Engineering*, *26*(1_suppl), S871–S879. https://doi.org/10.3233/BME-151379

Xie, W., Zhong, Z., Wang, X., Quan, L., Lu, Y., & Fu, S. (2025). The Impact of Event Preview on Pilot Attention Allocation During a Multitasking Test. In D. Harris & W.-C. Li (Eds.), *Engineering Psychology and Cognitive Ergonomics* (Vol. 15777, pp. 276–286). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-93721-7_19

Xin, K., & Li, Z. (2020). Visual working memory load does not affect the overall stimulus processing time in visual search. *Quarterly Journal of Experimental Psychology*, *73*(3), 330–343. https://doi.org/10.1177/1747021819881622

Yamaguchi, V., & Bortz, G. (2021). Memória operacional em música: Principais marcos teóricos e alguns dos desafios em seu estudo. *Revista Música Hodie*, *21*. https://doi.org/10.5216/mh.v21.67358

Yang, J.-F., Wang, P.-C., Liu, X.-Y., Bian, M.-C., Chen, L.-C., Lv, S.-Y., Tao, J.-F., Suo, G.-Y., Xuan, S.-Q., Li, R., Zhang, J.-W., Shu, C.-M., & Dou, Z. (2023). Analysis on causes of chemical industry accident from 2015 to 2020 in Chinese mainland: A complex network theory approach. *Journal of Loss Prevention in the Process Industries*, *83*, 105061. https://doi.org/10.1016/j.jlp.2023.105061

Yang, M., Lei, X., & Rau, P.-L. P. (2024). Commanding Consumers' Visual Attention: Enhancement and Attenuation. In P.-L. P. Rau (Ed.), *Cross-Cultural Design* (Vol. 14700, pp. 270–284). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-60901-5_19

Yilmaz, A. A. (2025). Critical Connections: Network Analysis of Human Errors in Aviation Accidents. *The International Journal of Aerospace Psychology*, *35*(4), 203–229. https://doi.org/10.1080/24721840.2025.2531741

Young, S. D., Daniels, T. S., Evans, E. T., de Haag, M. U., & Duan, P. P. (2013). *Understanding crew decision making in the presence of complexity—A flight simulation experiment*. Scopus. AIAA Infotech at Aerospace (I at A) Conference. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84883664676&partnerID=40&md5=521d6d6ac70c84ca6e77c26d6641b8af

Yuan, J., Jia, B., Zhang, C., Tian, L., Yi, H., & Wei, L. (2025). Pilot mental workload analysis in the A320 traffic pattern based on HRV features. *Frontiers in Neuroergonomics*, *6*, 1672492. https://doi.org/10.3389/fnrgo.2025.1672492

Yuan, X., Ng, K. K. H., Yiu, C. Y., & Li, Q. (2025). Optimising Pilot-Aircraft Interaction: A Low-Cost Projection-Enhanced Head-Up Display (PE-HUD) with Neuro-Ergonomic Validation. In C. Stephanidis, M. Antona, S. Ntoa, & G. Salvendy (Eds.), *HCI International 2025 Posters* (Vol. 2523, pp. 134–143). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-94153-5_13

Yue, R., Li, J., & Han, M. (2021). Aviation Accident Causation Analysis Based on Complex Network Theory. *Transactions of Nanjing University of Aeronautics and Astronautics*, *38*(4), 646–655. Scopus. https://doi.org/10.16356/j.1005-1120.2021.04.011

Zaw, L. T., & Suyatinov, S. I. (2022). Formation of Situational Awareness Based on Mental Models in the Controlling Complex Technical Objects. In A. G. Kravets, A. A.

Bolshakov, & M. Shcherbakov (Eds.), *Society 5.0: Human-Centered Society Challenges and Solutions* (Vol. 416, pp. 341–349). Springer International Publishing. https://doi.org/10.1007/978-3-030-95112-2_28

Zhang, X., & Mahadevan, S. (2021). Bayesian network modeling of accident investigation reports for aviation safety assessment. *Reliability Engineering & System Safety*, *209*, 107371. https://doi.org/10.1016/j.ress.2020.107371

Zheng, Y., Lu, Y., Wang, Z., Huang, D., & Fu, S. (2015). Developing a Measurement for Task Complexity in Flight. *Aerospace Medicine and Human Performance*, *86*(8), 698–704. https://doi.org/10.3357/AMHP.4189.2015

APPENDIX A. NODE NOMENCLATURE

Finding Code	Node Number	Description
106201220	A1	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Altitude
106201020	A2	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Airspeed
106101008	A3	Aircraft-Aircraft oper/perf/capability-Aircraft capability-Climb capability
106200020	A4	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-(general)
106204208	A5	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Angle of attack
106201211	A6	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Altitude
106203508	A7	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Climb rate
106204220	A8	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Angle of attack
106204020	A9	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Descent/approach/glide path
106205220	A10	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Prop/rotor parameters
106203520	A11	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Climb rate
106201221	A12	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Altitude
106202020	A13	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Directional control
102344410	A14	Aircraft-Aircraft systems-Navigation system-Ground proximity system
106101021	A15	Aircraft-Aircraft oper/perf/capability-Aircraft capability-Climb capability
106202320	A16	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Lateral/bank control
106201021	A17	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Airspeed
106203521	A18	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Climb rate
106103508	A19	Aircraft-Aircraft oper/perf/capability-Aircraft capability-Maximum weight
106201011	A20	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Airspeed
102302010	A21	Aircraft-Aircraft systems-Ice/rain protection system-Intake anti-ice, deice
105853001	A22	Aircraft-Aircraft power plant-Engine (reciprocating)-Recip eng cyl section
105725005	A23	Aircraft-Aircraft power plant-Engine (turbine/turboprop)-Turbine section
106100008	A24	Aircraft-Aircraft oper/perf/capability-Aircraft capability-(general)

106103599	A25	Aircraft-Aircraft oper/perf/capability-Aircraft capability-Maximum weight
105742106	A26	Aircraft-Aircraft power plant-Ignition system-Spark plugs/igniters
106204520	A27	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Surface speed/braking
106203720	A28	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Descent rate
102282201	A29	Aircraft-Aircraft systems-Fuel system-Fuel pumps
102324208	A30	Aircraft-Aircraft systems-Landing gear system-Brake
102344417	A31	Aircraft-Aircraft systems-Navigation system-Ground proximity system
102344408	A32	Aircraft-Aircraft systems-Navigation system-Ground proximity system
102344407	A33	Aircraft-Aircraft systems-Navigation system
102313027	A34	Aircraft-Aircraft systems-Indicating/recording systems-Data recorders (flight/maint)
105000011	A35	Aircraft-Aircraft power plant-(general)-(general)
107101023	A36	Aircraft-Fluids/misc hardware-Fluids-Fuel
106201711	A37	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Configuration
106204711	A38	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Heading/course
106204011	A39	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Descent/approach/glide path
101150011	A40	Aircraft-Aircraft handling/service-Loading-(general)
105859705	A41	Aircraft-Aircraft power plant-Engine (reciprocating)-Recip eng wiring
105742105	A42	Aircraft-Aircraft power plant-Ignition system-Spark plugs/igniters
106203721	A43	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Descent rate
102280001	A44	Aircraft-Aircraft systems-Fuel system
107140016	A45	Aircraft-Fluids/misc hardware-Misc hardware-(general)
106204221	A46	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Angle of attack
106200008	A47	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-(general)
102230010	A48	Aircraft-Aircraft systems-Communications system-(general)
102290008	A49	Aircraft-Aircraft systems-Hydraulic power system-(general)
104631005	A50	Aircraft-Aircraft propeller/rotor-Main rotor drive-Engine/transmission coupling
105852006	A51	Aircraft-Aircraft power plant-Engine (reciprocating)-Recip engine power section

105852001	A52	Aircraft-Aircraft power plant-Engine (reciprocating)-Recip engine power section
107101024	A53	Aircraft-Fluids/misc hardware-Fluids-Fuel
102256299	A54	Aircraft-Aircraft systems-Equipment/furnishings-Emergency locator beacon
102344499	A55	Aircraft-Aircraft systems-Navigation system-Ground proximity system
106103511	A56	Aircraft-Aircraft oper/perf/capability-Aircraft capability-Maximum weight
105850005	A57	Aircraft-Aircraft power plant-Engine (reciprocating)-(general)
102240008	A58	Aircraft-Aircraft systems-Electrical power system-(general)
106204008	A59	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Descent/approach/glide path
102323012	A60	Aircraft-Aircraft systems-Landing gear system-Gear extension and retract sys
105853006	A61	Aircraft-Aircraft power plant-Engine (reciprocating)-Recip eng cyl section
106104008	A62	Aircraft-Aircraft oper/perf/capability-Aircraft capability-CG/weight distribution
104621099	A63	Aircraft-Aircraft propeller/rotor-Main rotor system-Main rotor blade system
105732210	A64	Aircraft-Aircraft power plant-Engine fuel and control-Fuel control/carburetor
204152044	B1	Personnel issues-Action/decision-Info processing/decision-Decision making/judgment
206304044	B2	Personnel issues-Task performance-Use of equip/info-Aircraft control
202154544	B3	Personnel issues-Psychological-Attention/monitoring-Monitoring environment
206102544	B4	Personnel issues-Task performance-Planning/preparation-Flight planning/navigation
204101544	B5	Personnel issues-Action/decision-Action-Incorrect action performance
206102044	B6	Personnel issues-Task performance-Planning/preparation-Weather planning
204101044	B7	Personnel issues-Action/decision-Action-Incorrect action selection
202203544	B8	Personnel issues-Psychological-Perception/orientation/illusion-Situational awareness
204152040	B9	Personnel issues-Action/decision-Info processing/decision-Decision making/judgment
202202544	B10	Personnel issues-Psychological-Perception/orientation/illusio-Spatial disorientation
206301544	B11	Personnel issues-Task performance-Use of equip/info-Use of equip/system
203103044	B12	Personnel issues-Experience/knowledge-Experience/qualifications-Total instrument experience
204151044	B13	Personnel issues-Action/decision-Info processing/decision-Identification/recognition
201302044	B14	Personnel issues-Physical-Health/Fitness-Use of medication/drugs

204102540	B15	Personnel issues-Action/decision-Action-Delayed action
206300044	B16	Personnel issues-Task performance-Use of equip/info-(general)
204103044	B17	Personnel issues-Action/decision-Action-Lack of action
206000044	B18	Personnel issues-Task performance-(general)
203101044	B19	Personnel issues-Experience/knowledge-Experience/qualifications-Qualification/certification
206101044	B20	Personnel issues-Task performance-Planning/preparation-Performance calculations
202152044	B21	Personnel issues-Psychological-Attention/monitoring-Task monitoring/vigilance
204102544	B22	Personnel issues-Action/decision-Action-Delayed action
201205044	B23	Personnel issues-Physical-Impairment/incapacitation-Cardiovascular
202154540	B24	Personnel issues-Psychological-Attention/monitoring-Monitoring environment
204152036	B25	Personnel issues-Action/decision-Info processing/decision-Decision making/judgment
202153540	B26	Personnel issues-Psychological-Attention/monitoring-Monitoring other person
204100044	B27	Personnel issues-Action/decision-Action-(general)
202203044	B28	Personnel issues-Psychological-Perception/orientation/illusion-Geographic disorient (lost)
206102546	B29	Personnel issues-Task performance-Planning/preparation-Flight planning/navigation
202101044	B30	Personnel issues-Psychological-Personality/attitude-Self confidence
202101044	B30	Personnel issues-Psychological-Personality/attitude
206303544	B31	Personnel issues-Task performance-Use of equip/info-Use of available resources
206303544	B31	Personnel issues-Task performance-Use of equip/info
202201544	B32	Personnel issues-Psychological-Perception/orientation/illusion-Visual illusion/disorientation
203101544	B33	Personnel issues-Experience/knowledge-Experience/qualifications
201202544	B34	Personnel issues-Physical-Impairment/incapacitation-Prescription medication
206304040	B35	Personnel issues-Task performance-Use of equip/info-Aircraft control
201202044	B36	Personnel issues-Physical-Impairment/incapacitation-Illicit drug
204104544	B37	Personnel issues-Action/decision-Action-Unnecessary action
206304036	B38	Personnel issues-Task performance-Use of equip/info-Aircraft control
206201041	B39	Personnel issues-Task performance-Maintenance-Scheduled/routine maintenance

206303236	B40	Personnel issues-Task performance-Use of equip/info-Use of policy/procedure
206301540	B41	Personnel issues-Task performance-Use of equip/info-Use of equip/system
203202044	B42	Personnel issues-Experience/knowledge-Knowledge-Knowledge of meteorologic cond
204101036	B43	Personnel issues-Action/decision-Action-Incorrect action selection
204103036	B44	Personnel issues-Action/decision-Action-Lack of action
204152042	B45	Personnel issues-Action/decision-Info processing/decision
202102544	B46	Personnel issues-Psychological-Personality/attitude
204152046	B47	Personnel issues-Action/decision-Info processing/decision-Decision making/judgment
203201044	B48	Personnel issues-Experience/knowledge-Knowledge
206203042	B49	Personnel issues-Task performance-Maintenance-Modification/alteration
204103040	B50	Personnel issues-Action/decision-Action-Lack of action
202151044	B51	Personnel issues-Psychological-Attention/monitoring
206102536	B52	Personnel issues-Task performance-Planning/preparation-Flight planning/navigation
201151044	B53	Personnel issues-Physical-Sensory ability/limitation-Visual function
206200041	B54	Personnel issues-Task performance-Maintenance-(general)
206302544	B55	Personnel issues-Task performance-Use of equip/info-Use of charts
202154536	B56	Personnel issues-Psychological-Attention/monitoring-Monitoring environment
206351036	B57	Personnel issues-Task performance-Communication (personnel)-Lack of communication
204103031	B58	Personnel issues-Action/decision-Action-Lack of action
206203541	B59	Personnel issues-Task performance-Maintenance-Installation
204101541	B60	Personnel issues-Action/decision-Action-Incorrect action performance
204150044	B61	Personnel issues-Action/decision-Info processing/decision-(general)
202152040	B62	Personnel issues-Psychological-Attention/monitoring-Task monitoring/vigilance
202300044	B63	Personnel issues-Psychological-Cognitive limitation-(general)
201204544	B64	Personnel issues-Physical-Impairment/incapacitation-Neurological
201203044	B65	Personnel issues-Physical-Impairment/incapacitation-OTC medication
201203544	B66	Personnel issues-Physical-Impairment/incapacitation-Hypoxia/anoxia

202102044	B67	Personnel issues-Psychological-Personality/attitude-Complacency
201201544	B68	Personnel issues-Physical-Impairment/incapacitation-Alcohol
204104540	B69	Personnel issues-Action/decision-Action-Unneccessary action
204101536	B70	Personnel issues-Action/decision-Action-Incorrect action performance
206300036	B71	Personnel issues-Task performance-Use of equip/info-(general)
206100044	B72	Personnel issues-Task performance-Planning/preparation-(general)
203201544	B73	Personnel issues-Experience/knowledge-Knowledge-Knowledge of equipment
203103544	B74	Personnel issues-Experience/knowledge-Experience/qualifications-Recent experience
204151040	B75	Personnel issues-Action/decision-Info processing/decision-Identification/recognition
206102540	B76	Personnel issues-Task performance-Planning/preparation-Flight planning/navigation
206303540	B77	Personnel issues-Task performance-Use of equip/info-Use of available resources
203152544	B78	Personnel issues-Experience/knowledge-Training
204152544	B79	Personnel issues-Action/decision-Info processing/decision
202250044	B80	Personnel issues-Psychological-Mental/emotional state-(general)
203100044	B81	Personnel issues-Experience/knowledge-Experience/qualifications-(general)
201303544	B82	Personnel issues-Physical-Health/Fitness-Predisposing condition
204000044	B83	Personnel issues-Action/decision-(general)-(general)
206402044	B84	Personnel issues-Task performance-Workload management
203203040	B85	Personnel issues-Experience/knowledge-Knowledge-Knowledge of geographic area
206103044	B86	Personnel issues-Task performance-Planning/preparation-Fuel planning
204152031	B87	Personnel issues-Action/decision-Info processing/decision-Decision making/judgment
303507584	C1	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Below VFR minima
302101086	C2	Environmental issues-Physical environment-Terrain-Mountainous/hilly terrain
302101087	C3	Environmental issues-Physical environment-Terrain-Mountainous/hilly terrain
303507582	C4	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Below VFR minima
303501082	C5	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Low ceiling
303102082	C6	Environmental issues-Conditions/weather/phenomena-Temp/humidity/pressure-High density altitude

303602082	C7	Environmental issues-Conditions/weather/phenomena-Light condition-Dark
303403582	C8	Environmental issues-Conditions/weather/phenomena-Wind-Downdraft
302202091	C9	Environmental issues-Physical environment-Object/animal/substance-Tree(s)
303102081	C10	Environmental issues-Conditions/weather/phenomena-Temp/humidity/pressure-High density altitude
303201082	C11	Environmental issues-Conditions/weather/phenomena-Turbulence-Terrain induced turbulence
302101081	C12	Environmental issues-Physical environment-Terrain-Mountainous/hilly terrain
303502582	C13	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Clouds
303503082	C14	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Obscuration
303602091	C15	Environmental issues-Conditions/weather/phenomena-Light condition-Dark
303403586	C16	Environmental issues-Conditions/weather/phenomena-Wind-Downdraft
302202082	C17	Environmental issues-Physical environment-Object/animal/substance-Tree(s)
303602083	C18	Environmental issues-Conditions/weather/phenomena-Light condition-Dark
303403591	C19	Environmental issues-Conditions/weather/phenomena-Wind-Downdraft
303501582	C20	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Low visibility
303403581	C21	Environmental issues-Conditions/weather/phenomena-Wind-Downdraft
303501091	C22	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Low ceiling
303507591	C23	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Below VFR minima
303401582	C24	Environmental issues-Conditions/weather/phenomena-Wind-Tailwind
303501099	C25	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Low ceiling
303501591	C26	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Low visibility
303507585	C27	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Below VFR minima
303503091	C28	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Obscuration
303503084	C29	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Obscuration
303502599	C30	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Clouds
303502584	C31	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Clouds
303404582	C32	Environmental issues-Conditions/weather/phenomena-Wind-Gusts
303102091	C33	Environmental issues-Conditions/weather/phenomena-Temp/humidity/pressure-High density altitude

303401091	C34	Environmental issues-Conditions/weather/phenomena-Wind-Sudden wind shift
303402582	C35	Environmental issues-Conditions/weather/phenomena-Wind-Variable wind
303505582	C36	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Fog
303502591	C37	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Clouds
303000082	C38	Environmental issues-Conditions/weather/phenomena-(general)-(general)
303603083	C39	Environmental issues-Conditions/weather/phenomena-Light condition-Glare
303104082	C40	Environmental issues-Conditions/weather/phenomena-Temp/humidity/pressure-Thermal lifting
303401591	C41	Environmental issues-Conditions/weather/phenomena-Wind-Tailwind
303503591	C42	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Rain
303501599	C43	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Low visibility
303404581	C44	Environmental issues-Conditions/weather/phenomena-Wind-Gusts
302202085	C45	Environmental issues-Physical environment-Object/animal/substance-Tree(s)
303602099	C46	Environmental issues-Conditions/weather/phenomena-Light condition-Dark
303501584	C47	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Low visibility
303501084	C48	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Low ceiling
301601099	C49	Environmental issues-Operating environment-Communication system-VHF/HF radio
303503085	C50	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Obscuration
302102584	C51	Environmental issues-Physical environment-Terrain-High elevation
303403584	C52	Environmental issues-Conditions/weather/phenomena-Wind-Downdraft
303504084	C53	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Freezing rain/sleet
301301099	C54	Environmental issues-Operating environment-Meteorological services-Meteo equip coverage/avail
303404081	C55	Environmental issues-Conditions/weather/phenomena-Wind-Crosswind
302100099	C56	Environmental issues-Physical environment-Terrain-(general)
303600099	C57	Environmental issues-Conditions/weather/phenomena-Light condition-(general)
303503083	C58	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Obscuration
303103082	C59	Environmental issues-Conditions/weather/phenomena-Temp/humidity/pressure-Conducive to carburetor icing
303103081	C60	Environmental issues-Conditions/weather/phenomena-Temp/humidity/pressure-Conducive to carburetor icing

302100091	C61	Environmental issues-Physical environment-Terrain-(general)
303404591	C62	Environmental issues-Conditions/weather/phenomena-Wind-Gusts
302102582	C63	Environmental issues-Physical environment-Terrain-High elevation
303600082	C64	Environmental issues-Conditions/weather/phenomena-Light condition-(general)
303502091	C65	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Haze/smoke
303404082	C66	Environmental issues-Conditions/weather/phenomena-Wind-Crosswind
302205585	C67	Environmental issues-Physical environment-Object/animal/substance-Debris/dirt/foreign object
302102591	C68	Environmental issues-Physical environment-Terrain-High elevation
303102084	C69	Environmental issues-Conditions/weather/phenomena-Temp/humidity/pressure-High density altitude
303200082	C70	Environmental issues-Conditions/weather/phenomena-Turbulence-(general)
303102085	C71	Environmental issues-Conditions/weather/phenomena-Temp/humidity/pressure-High density altitude
303500099	C72	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-(general)
302300084	C73	Environmental issues-Physical environment-Runway/land/takeoff/taxi surfa-(general)
302302091	C74	Environmental issues-Physical environment-Runway/land/takeoff/taxi surfa-Soft
303505591	C75	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Fog
303502583	C76	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Clouds
303401581	C77	Environmental issues-Conditions/weather/phenomena-Wind-Tailwind
303500091	C78	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-(general)
303402591	C79	Environmental issues-Conditions/weather/phenomena-Wind-Variable wind
302103082	C80	Environmental issues-Physical environment-Terrain-Snowy/icy
303602582	C81	Environmental issues-Conditions/weather/phenomena-Light condition-Flat light
303502099	C82	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Haze/smoke
303504582	C83	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Snow
302103081	C84	Environmental issues-Physical environment-Terrain-Snowy/icy terrain
303602583	C85	Environmental issues-Conditions/weather/phenomena-Light condition-Flat light
303403082	C86	Environmental issues-Conditions/weather/phenomena-Wind-Updraft
303403599	C87	Environmental issues-Conditions/weather/phenomena-Wind-Downdraft

302202099	C88	Environmental issues-Physical environment-Object/animal/substance-Tree(s)
303401592	C89	Environmental issues-Conditions/weather/phenomena-Wind-Tailwind
303200099	C90	Environmental issues-Conditions/weather/phenomena-Turbulence-(general)
303404586	C91	Environmental issues-Conditions/weather/phenomena-Wind-Gusts
302202081	C92	Environmental issues-Physical environment-Object/animal/substance-Tree(s)
303507583	C93	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Below VFR minima
303400082	C94	Environmental issues-Conditions/weather/phenomena-Wind-(general)
302103091	C95	Environmental issues-Physical environment-Terrain-Snowy/icy
303506082	C96	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Whiteout
302201791	C97	Environmental issues-Physical environment-Object/animal/substance-Tower/antenna (incl guy wires)
303601091	C98	Environmental issues-Conditions/weather/phenomena-Light condition-Bright light
302103099	C99	Environmental issues-Physical environment-Terrain-Snowy/icy
302200082	C100	Environmental issues-Physical environment-Object/animal/substance-(general)
302203282	C101	Environmental issues-Physical environment-Object/animal/substance-Wire
303501589	C102	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Low visibility
303501585	C103	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Low visibility
303403588	C104	Environmental issues-Conditions/weather/phenomena-Wind-Downdraft
302202582	C105	Environmental issues-Physical environment-Object/animal/substance-Ground vehicle
303000086	C106	Environmental issues-Conditions/weather/phenomena-(general)-(general)
302102581	C107	Environmental issues-Physical environment-Terrain-High elevation
302300099	C108	Environmental issues-Physical environment-Runway/land/takeoff/taxi surfa-(general)
303502585	C109	Environmental issues-Conditions/weather/phenomena-Ceiling/visibility/precip-Clouds
301402092	C110	Environmental issues-Operating environment-Air traffic/operating proc-ATC clearance procedure
302205291	C111	Environmental issues-Physical environment-Object/animal/substance-Hidden/submerged object
302101591	C112	Environmental issues-Physical environment-Terrain-Rough terrain
106200021	A65	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-(general)
106000008	A66	Aircraft-Aircraft oper/perf/capability-(general)-(general)

106105008	A67	Aircraft-Aircraft oper/perf/capability-Aircraft capability-Braking capability
102275011	A68	Aircraft-Aircraft systems-Flight control system-TE flap control system
106202420	A69	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Yaw control
106202220	A70	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Pitch control
104641005	A71	Aircraft-Aircraft propeller/rotor-Tail rotor-Tail rotor blade
106202408	A72	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Yaw control
106205211	A73	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Prop/rotor parameters
105710001	A74	Aircraft-Aircraft power plant-Power plant-(general)
106204720	A75	Aircraft-Aircraft oper/perf/capability-Performance/control parameters-Heading/course
105852013	A76	Aircraft-Aircraft power plant-Engine (reciprocating)-Recip engine power section
102282113	A77	Aircraft-Aircraft systems-Fuel system-Fuel filter-strainer
105732217	A78	Aircraft-Aircraft power plant-Engine fuel and control-Fuel control/carburetor
107101025	A79	Aircraft-Fluids/misc hardware-Fluids-Fuel
105850099	A80	Aircraft-Aircraft power plant-Engine (reciprocating)-(general)
104671001	A81	Aircraft-Aircraft propeller/rotor-Rotorcraft flight control-Main rotor control
105741401	A82	Aircraft-Aircraft power plant-Ignition system-Magneto/distributor
104652099	A83	Aircraft-Aircraft propeller/rotor-Tail rotor drive system-Tail rotor gearbox

APPENDIX B. NASA-TASK LOAD INDEX (NASA-TLX)

NASA-TLX Mental Workload Rating Scale

Subject ID: ____________

Please place the slider bar along the scale at the point that best indicates your experience with the approach. The scale is divided into 20 equal intervals anchored by bipolar descriptors (High/Low). The 21 vertical tick marks on each scale divide the scale from 0 to 100 increments of 5.

Mental Demand: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the mission easy or demanding, simple or complex, exacting or forgiving?

Low High

Physical Demand: How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the mission easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low High

Temporal Demand: How much time pressure did you feel due to the rate or pace at which the mission occurred? Was the pace slow and leisurely or rapid and frantic?

Low High

Performance: How successful do you think you were in accomplishing the goals of the mission? How satisfied were you with your performance in accomplishing these goals?

Low High

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low High

Frustration: How discouraged, stressed, irritated, and annoyed versus gratified, relaxed, content, and complacent did you feel during your mission?

Low High

Total Un-weighted Mental Workload Score:____________

NASA-TLX Mental Workload Rankings

For each pair listed below, circle the scale title that represents the most important contributor to workload in the display.

	Subscales	
Mental Demand	or	Physical Demand
Mental Demand	or	Temporal Demand
Mental Demand	or	Performance
Mental Demand	or	Effort
Mental Demand	or	Frustration
Physical Demand	or	Temporal Demand
Physical Demand	or	Performance
Physical Demand	or	Effort
Physical Demand	or	Frustration
Temporal Demand	or	Performance
Temporal Demand	or	Frustration
Temporal Demand	or	Effort
Performance	or	Frustration
Performance	or	Effort
Frustration	or	Effort

NASA-TLX Mental Workload Rankings

Please assign points (0-100) to each subscale based on their relative importance. The **most important gets 100 points**, and others are scored relative to it. You can assign equal values wherever necessary.

Sub-Scales	Score (0-100)
Mental Demands	
Physical Demands	
Temporal Demands	
Performance	
Effort	
Frustration	
Total Score	

APPENDIX C. SITUATIONAL AWARENESS RATING TECHNIQUE (SART)

For each of the 10 contributing factors, please mark one of the seven points that best represents your experience.

SART Items

1. Instability of Situation

How changeable is the situation? Is the situation highly unstable and likely to change suddenly (High), or is it very stable and straightforward (Low)?

1	2	3	4	5	6	7
Low						High

2. Complexity of Situation

How complicated is the situation? Is it complex with many interrelated components (High), or is it simple and straightforward (Low)?

1	2	3	4	5	6	7
Low						High

3. Variability of Situation

How many variables are changing within the situation? Are there a large number of factors varying (High), or are there very few variables changing (Low)?

1	2	3	4	5	6	7
Low						High

4. Arousal

How aroused are you in the situation? Are you alert and ready for activity (High), or do you have a low degree of alertness (Low)?

1	2	3	4	5	6	7
Low						High

5. Concentration of Attention

How much are you concentrating on the situation? Are you concentrating on many aspects of the situation (High), or focused on only one (Low)?

1	2	3	4	5	6	7
Low						High

6. Division of Attention

How much is your attention divided in the situation? Are you concentrating on many aspects of the situation (High), or focused on only one (Low)?

1	2	3	4	5	6	7
Low						High

7. Spare Mental Capacity

How much mental capacity do you have to spare in the situation? Do you have sufficient capacity to attend to many variables (High), or nothing to spare at all (Low)?

1	2	3	4	5	6	7
Low						High

8. Information Quantity

How much information have you gained about the situation? Have you received and understood a great deal of knowledge (High), or very little (Low)?

1	2	3	4	5	6	7
Low						High

9. Information Quality

How good is the quality of the information you have gathered about the situation? Is the information available very useful (High), or not usable at all (Low)?

1	2	3	4	5	6	7
Low						High

10. Familiarity with Situation

How familiar are you with the situation? Do you have a great deal of relevant experience (High), or is it a new situation (Low)?

1	2	3	4	5	6	7
Low						High

Additional Control-Mode Question

Please select which of the following statements applies most to you in the situation:

Options	**Response**
a)	You do not fully understand the situation; some events surprise you, and your actions are reactions to events without a plan.
b)	You understand the situation at the moment and base your decisions on the current picture with very limited planning ahead.
c)	You have some understanding of the situation and can plan your actions ahead for a short period of time.
d)	You have a good understanding of what is happening and can plan ahead for most of the scenario.